



Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases

Birte Zurek¹ · Kornelia Ellwanger¹ · Lisenka E. L. M. Vissers^{2,3} · Rebecca Schüle^{4,5} · Matthis Synofzik^{4,5} · Ana Töpfer⁶ · Richarda M. de Voer^{2,7} · Steven Laurie⁸ · Leslie Matalonga⁸ · Christian Gilissen^{2,7} · Stephan Ossowski¹ · Peter A. C. 't Hoen^{7,9} · Antonio Vitobello¹⁰ · Julia M. Schulze-Hentrich¹ · Olaf Riess^{1,11} · Han G. Brunner^{2,3,12} · Anthony J. Brookes¹³ · Ana Rath¹⁴ · Gisèle Bonne¹⁵ · Gulcin Gumus¹⁶ · Alain Verloes¹⁷ · Nicoline Hoogerbrugge^{2,7} · Teresinha Evangelista¹⁵ · Tina Harmuth¹ · Morris Swertz¹⁸ · Dylan Spalding¹⁹ · Alexander Hoischen^{2,7,20} · Sergi Beltran^{8,21,22} · Holm Graessner^{1,11} · Solve-RD consortium

Received: 14 October 2020 / Revised: 8 February 2021 / Accepted: 4 March 2021
© The Author(s) 2021. This article is published with open access

Abstract

For the first time in Europe hundreds of rare disease (RD) experts team up to actively share and jointly analyse existing patient's data. Solve-RD is a Horizon 2020-supported EU flagship project bringing together >300 clinicians, scientists, and patient representatives of 51 sites from 15 countries. Solve-RD is built upon a core group of four European Reference Networks (ERNs; ERN-ITHACA, ERN-RND, ERN-Euro NMD, ERN-GENTURIS) which annually see more than 270,000 RD patients with respective pathologies. The main ambition is to solve unsolved rare diseases for which a molecular cause is not yet known. This is achieved through an innovative clinical research environment that introduces novel ways to organise expertise and data. Two major approaches are being pursued (i) massive data re-analysis of >19,000 unsolved rare disease patients and (ii) novel combined -omics approaches. The minimum requirement to be eligible for the analysis activities is an inconclusive exome that can be shared with controlled access. The first preliminary data re-analysis has already diagnosed 255 cases from 8393 exomes/genome datasets. This unprecedented degree of collaboration focused on sharing of data and expertise shall identify many new disease genes and enable diagnosis of many so far undiagnosed patients from all over Europe.

Rare Diseases (RD) are individually rare but collectively a common health issue. Around 80% of RD are estimated to have a genetic cause [1]. The time to a genetic diagnosis however often takes several years and initial clinical diagnoses are incorrect in up to 40% of families [2]. Around 50% of patients with a RD remain undiagnosed even in advanced expert clinical settings where whole exome sequencing (WES) is applied routinely as a diagnostic approach. Depending on the exact diagnostic setting, the

inclusion criteria and the type of RD, the diagnostic yield from WES ranges between 15 and 51% of cases [3, 4].

At least two scenarios allow boosting the current yield of WES. Firstly, there is a value in re-analysing WES data regularly [5] and on massive scale [6], but not every RD expert has access to tools enabling this systematically. Secondly, it is clear that moving beyond the exome can provide additional benefits [7, 8].

Solve-RD aims to solve a large number of unsolved RD, for which a molecular cause is not yet known, by implementing both strategies mentioned above. To this end, Solve-RD applies innovative ways to effectively organise expertise and data.

Members of the Solve-RD consortium are listed below
Acknowledgements.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-021-00859-0>.

✉ Holm Graessner
holm.graessner@med.uni-tuebingen.de

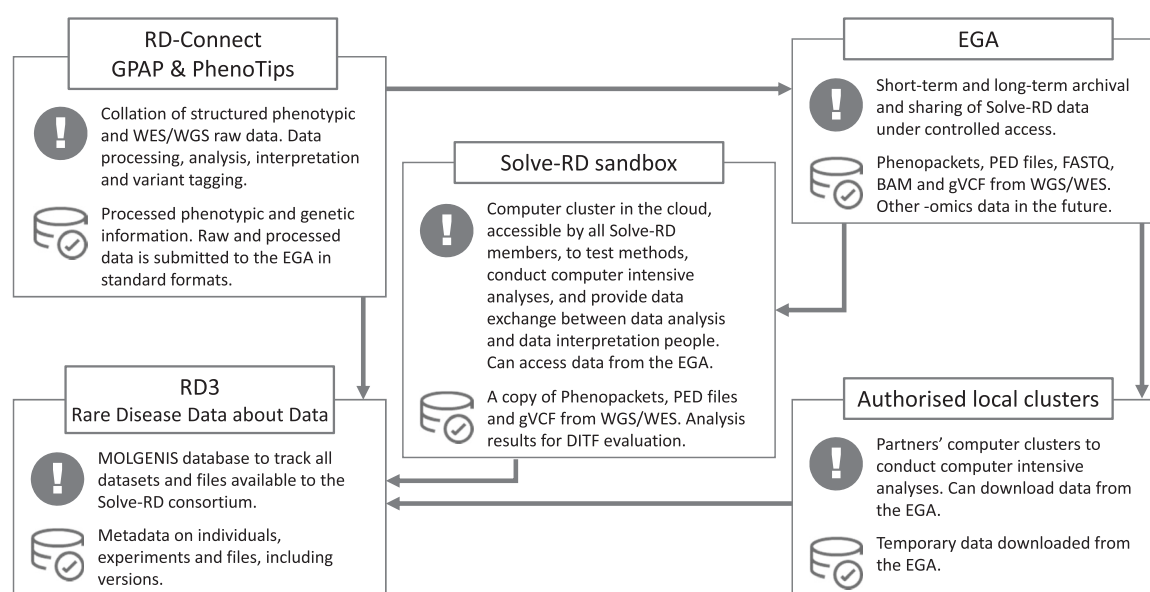
Extended author information available on the last page of the article.

Cohorts

To structure its work Solve-RD has defined four types of *cohorts*. *Cohort 1*, “Unsolved Cases”, comprises cases with an inconclusive WES or whole genome sequencing (WGS)

Table 1 Examples for the specific ERN cohorts and the unsolvables.

Cohort	Rationale
<i>Cohort 2: Long-read whole genome sequencing (LR-WGS)</i>	
X-linked spinal and bulbar muscular atrophy (SBMA)	Suspected expansions of repeat disorder or other hidden structural variants (SV)
Hereditary ataxia	Suspected expansions of repeat disorder or other hidden SVs
<i>Cohort 2: Genomics and Epigenomics</i>	
Unexplained Intellectual Disability (ID): patient-parent trios	De novo mutation prioritisation very powerful filter for de novo methylation changes
Diffuse gastric cancer	Hypermethylation of cancer gene promoter known disease mechanism
Rare pheochromocytomas and paragangliomas	Hypermethylation of cancer gene promoter known disease mechanism
<i>Cohort 4</i>	
Unsolved syndromes available via ERN ITHACA	Aicardi syndrome, Gomez–Lopez Hernandez syndrome, Hallermann–Streiff syndrome are clinically well-defined entities and have been studied by WES and WGS globally and remain unsolved

**Fig. 1 Solve-RD data infrastructure.** Key components of the Solve-RD infrastructure for multi-omics data analysis, illustrating main use and data available.

from any partnering or associated ERN center. These data undergo a comprehensive re-analysis effort. *Cohort 2*, “Specific ERN Cohorts”, represent disease group specific ERN cohorts that are analysed by newly applied tailored -omics approaches. *Cohort 3*, “Ultra-Rare Rare Diseases”, includes (groups of) patients with unique phenotypes identified (and matched) by RD experts from all ERN participants. For the diseases included in *Cohort 4*, “The Unsolvables”, all relevant -omics methodologies will be used to solve highly recognisable, clinically well-defined disease entities for which the disease cause has not been found yet despite considerable previous research investigations including WES and WGS (Table 1).

In total, Solve-RD is targeting to re-analyse >19,000 datasets for cohort 1, sequence ~3500 short- and long-read

WGS for cohorts 2, 3, and 4 and add >3500 additional -omics experiments including RNA sequencing, epigenomics, metabolomics, Deep-WES, and deep molecular phenotyping. Data collected and produced in Solve-RD shall be shared via the European Genome-Phenome Archive (EGA) and the RD-Connect Genome-Phenome Analysis Platform (GPAP) to allow controlled access by other RD initiatives and scientists.

Organisation of data

The Solve-RD strategy relies on the availability of large amounts of good quality, standardised genomic and phenotypic data and metadata from undiagnosed RD

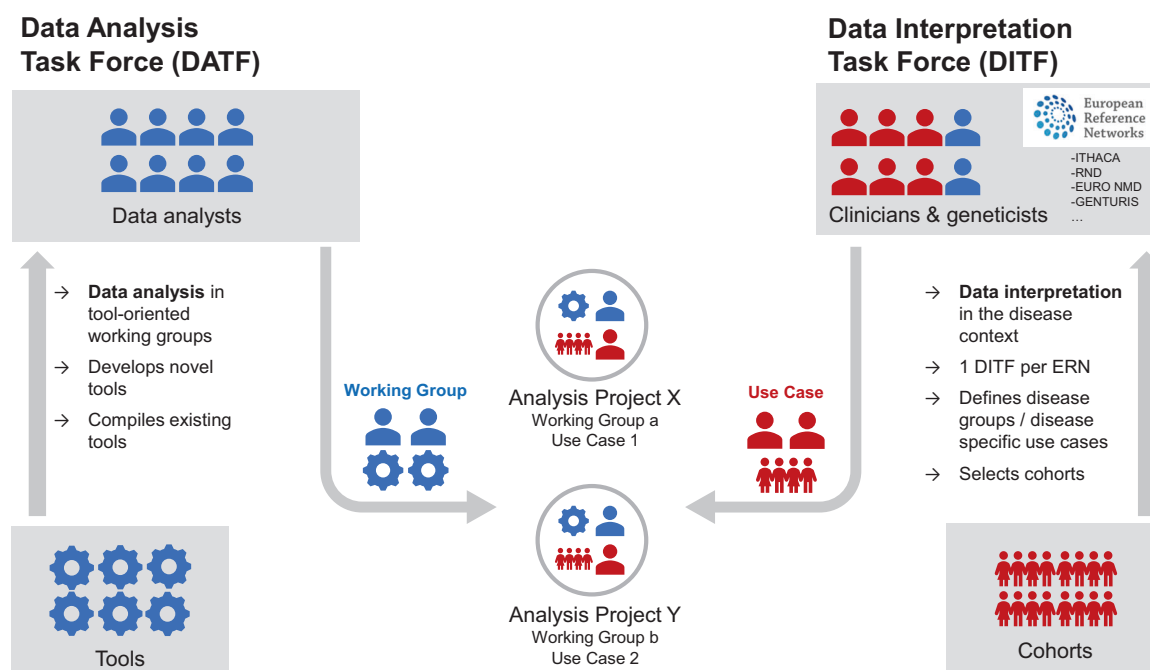


Fig. 2 The Solve-RD data analysis structure ‘in action’. Consisting of the Data Analysis Task Force (DATF) and four Data Interpretation Task Forces (DITF)—one per core ERN involved. The DATF

established working groups (WGs) for specific analyses. Working groups and DITFs jointly work on analysis projects based on use cases described by the DITF members.

patients and their relatives. Solve-RD follows a centralised approach, to enable all envisioned analyses. Data sharing in Solve-RD is regulated by policy documents, available on the project’s website. To overcome the technical challenge of centralising large amounts of data, Solve-RD leverages existing infrastructures such as EGA, GPAP, and computing clusters from project partners (Fig. 1). In addition, Solve-RD is developing a cloud-based computing cluster for collaborative analysis and methods testing (the Solve-RD Sandbox) and a central database to control and view all the project’s data and metadata (RD3; rare disease data about data) using the MOLGENIS open source data platform [9]. Clinical data and pedigree structure for all participating individuals is collated through standard terms and ontologies such as HPO, ORDO, and OMIM using GPAP-PhenoStore. To share data within the project and beyond, Solve-RD is an early adopter of the recently GA4GH-approved (Global Alliance for Genomics and Health, <https://www.ga4gh.org>) PhenoPackets standard to enable exchange of phenotypic and family information [10].

For each individual, WES and/or WGS data are submitted to GPAP in FASTQ, BAM, or CRAM format. The sequencing data are processed through a standard pipeline based on GATK (Genomic Analysis Toolkit variant calling software) best practices [11, 12]. After that, PhenoPackets, PED files (for pedigrees), raw data (FASTQ), alignments (BAM) and genetic variants (gVCF) are transferred to the EGA, where they are archived and made available to the consortium (and later on to the broader RD community) for

further analysis. Furthermore, Solve-RD data are connected to MatchMaker Exchange via GPAP.

To reach the ambitious goal to collect 19,000 unsolved WES/WGS, Solve-RD has defined several deadlines to submit data to the project. After each deadline, all data are processed and released as a data freeze, which is amenable to corrections via patches. The first data freeze, released in early 2020, includes data from 8,393 individuals.

In parallel to the collection of existing data for cohort 1, new omics data are being generated for cohorts 2, 3, and 4. A common data workflow has been established for all these data types (Fig. 1). The data collated and generated by Solve-RD constitutes a unique collection that will be valuable beyond the project, and the consortium is committed to make it FAIR under controlled access, through the EGA and GPAP.

Organisation of expertise

Solve-RD works on the interphase of many disciplines relevant to solving the unsolved RD. Central to the RD field are clinical geneticists and clinical scientists organised in the respective ERNs. Solve-RD provides expertise in genomics and other -omics data analysis, through data scientists, molecular geneticists, and bioinformaticians.

To warrant the best exchange of expertise we have implemented two structures: (i) Data scientists and genomics experts are organised in a Data Analysis Task Force

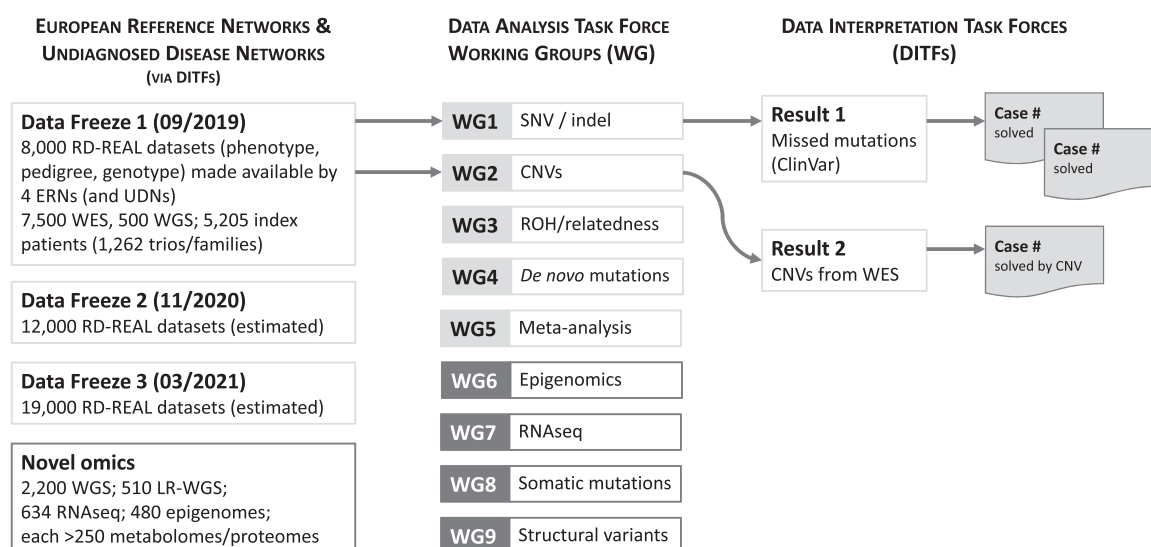


Fig. 3 Organisation of new result flow in Solve-RD. Working groups (WG) 1–5 will re-analyse existing sequencing data. Novel

omics data will be analysed by all working groups (as appropriate). RD-REAL refers to Rare Disease - REAnalysis Logistics.

(DATF), (ii) Expert clinicians and geneticists from each ERN are organised in a Data Interpretation Task Force (DITF) (Fig. 2). The tasks for these structures are in brief: ►DITF: define needs of ERN for (a) data re-analysis and (b) novel -omics data; define use cases for re-analysis and novel analysis; discuss/test suitable data output formats for clinical scientists; coordinate collaborative data interpretation; discuss within respective ERN network and feedback to DATF. ►DATF: map expertise in Solve-RD and all (ERN-)partners; create *Analysis Projects* (Supplementary Table S1) based on ERNs needs; develop state-of-the-art analysis tools; analyse data: (a) data re-analysis and (b) novel omics data; optimise data sharing and output formats for DITF/ERNs.

The structure implemented for data re-analysis has proven efficient and versatile [13], and will therefore be applied for novel omics data analysis, with additional working groups for specific -omics technologies (Fig. 3).

To integrate expertise not available within the Solve-RD consortium, particularly with regards to molecular and functional validation of newly found genes, Solve-RD is implementing an innovative brokerage system (Rare Disease Models and Mechanisms Network—Europe (RDMM-Europe)) that has already been successfully used in Canada [14]. As of 4 December 2020, 14 “brokering” Seeding Grants have been awarded to external model investigators.

Achievements and challenges

The work of the first 3 years of Solve-RD resulted in a practical solution to share and jointly analyse 8393 datasets

from all over Europe: Solve-RD organised RD expertise via a DITF and DATF with the respective working group structure described above. The first re-analysis approaches resulted in 255 newly diagnosed cases, mainly by leveraging latest ClinVar entries. As examples we refer to adjacent articles, published jointly in this issue [13, 15–18]. Many more candidate variants and new analysis results are under evaluation.

To achieve its current status Solve-RD has successfully addressed some critical challenges that are (a) European data sharing in accordance with GDPR, (b) heterogeneity in existing WES data (e.g. 26 WES kits so far; multiple sequencing platforms), (c) implementing a centralised analysis approach and (d) addressing the rarity of events.

It is the vision of Solve-RD that, by the end of the project, the Solve-RD dataset will be the largest well-annotated, standardised, multi-omics RD dataset on the diseases covered by the four core ERNs. In this sense, we hope that the Solve-RD dataset will be as useful to the RD community as the gnomAD consortium is for the genomics community [19], by making -omics data of RD populations available to the community.

Acknowledgements The Solve-RD project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 779257. This research is supported (not financially) by four ERNs: (1) The ERN for Intellectual Disability, Telehealth and Congenital Anomalies (ERN-ITHACA)—Project ID No 869189; (2) The ERN on Rare Neurological Diseases (ERN-RND)—Project ID No 739510; (3) The ERN for Neuromuscular Diseases (ERN Euro-NMD)—Project ID No 870177; (4) The ERN on Genetic Tumour Risk Syndromes (ERN GENTURIS)—Project ID No 739547. The ERNs are co-funded by the European Union within the framework of the Third Health Programme.

Solve-RD consortium Olaf Riess^{1,11}, Tobias B. Haack¹, Holm Graessner^{1,11}, Birte Zurek^{1,11}, Kornelia Ellwanger^{1,11}, Stephan Ossowski¹, German Demidov¹, Marc Sturm¹, Julia M. Schulze-Hentrich¹, Rebecca Schüle^{4,5}, Christoph Kessler^{4,5}, Melanie Wayand^{4,5}, Matthias Synofzik^{4,5}, Carlo Wilke^{4,5}, Andreas Traschütz^{4,5}, Ludger Schöls^{4,5}, Holger Hengel^{4,5}, Peter Heutink^{4,5}, Han Brunner^{2,3,12}, Hans Scheffer^{2,12}, Nicoline Hoogerbrugge^{2,7}, Alexander Hoischen^{2,7,20}, Peter A.C. 't Hoen^{7,9}, Lisenka E.L.M. Vissers^{2,3}, Christian Gilissen^{2,7}, Wouter Steyaert^{2,7}, Karolis Sablauskas², Richarda M. de Voer^{2,7}, Erik-Jan Kamsteeg², Bart van de Warrenburg^{3,23}, Nienke van Os^{3,23}, Iris te Paske^{2,7}, Erik Janssen^{2,7}, Elke de Boer^{2,3}, Marloes Stehouwer², Burcu Yaldiz², Tjitske Kleefstra^{2,3}, Anthony J. Brookes¹³, Colin Veal¹³, Spencer Gibson¹³, Marc Wadswley¹³, Mehdi Mehtarizadeh¹³, Umar Riaz¹³, Greg Warren¹³, Farid Yavari Dizjikan¹³, Thomas Shorter¹³, Ana Töpf⁶, Volker Straub⁶, Chiara Marini Bettolo⁶, Sabine Specht⁶, Jill Clayton-Smith²⁴, Siddharth Banka^{24,25}, Elizabeth Alexander²⁴, Adam Jackson²⁴, Laurence Faivre^{10,26,27,28,29}, Christel Thauvin^{10,27,28,29}, Antonio Vitobello¹⁰, Anne-Sophie Denommé-Pichon¹⁰, Yannis Duffourd^{10,28}, Emilie Tisserant¹⁰, Ange-Line Bruel¹⁰, Christine Peyron^{30,31}, Aurore Pélissier³¹, Sergi Beltran^{8,21}, Ivo Glynne Gut²¹, Steven Laurie²¹, Davide Piscia²¹, Leslie Matalonga²¹, Anastasios Papakonstantinou²¹, Gemma Bullich²¹, Alberto Corvo²¹, Carles Garcia²¹, Marcos Fernandez-Callejo²¹, Carles Hernández²¹, Daniel Pico²¹, Ida Paramonov²¹, Hanns Lochmüller²¹, Gulcin Gumus³², Virginie Bros-Facer³³, Ana Rath¹⁴, Marc Hanauer¹⁴, Annie Olry¹⁴, David Lagorce¹⁴, Svitlana Havrylenko¹⁴, Katia Izem¹⁴, Fanny Rigour¹⁴, Giovanni Stevanin^{34,35,36,37,38}, Alexandra Durr^{35,36,37,39}, Claire-Sophie Davoine^{35,36,37,38}, Léna Guillot-Noel^{35,36,37,38}, Anna Heinzmann^{35,36,37,40}, Giulia Coarelli^{35,36,37,40}, Gisèle Bonne¹⁵, Teresinha Evangelista¹⁵, Valérie Allamand¹⁵, Isabelle Nelson¹⁵, Rabah Ben Yaou^{15,41,42}, Corinne Metay^{15,43}, Bruno Eymard^{15,41}, Enzo Cohen¹⁵, Antonio Atalaia¹⁵, Tanya Stojkovic^{15,41}, Milan Macek Jr.⁴⁴, Marek Turnovec⁴⁴, Dana Thomasová⁴⁴, Radka Pourová Kremlíková⁴⁴, Vera Franková⁴⁴, Markéta Havlovicová⁴⁴, Vlastimil Kremlík⁴⁴, Helen Parkinson¹⁹, Thomas Keane¹⁹, Dylan Spalding¹⁹, Alexander Senf¹⁹, Peter Robinson⁴⁵, Daniel Danis⁴⁵, Glenn Robert⁴⁶, Alessia Costa⁴⁶, Christine Patch^{46,47}, Mike Hanna⁴⁸, Henry Houlden⁴⁹, Mary Reilly⁴⁸, Jana Vandrovcova⁴⁹, Francesco Muntoni^{50,51}, Irina Zaharieva⁵⁰, Anna Sarkozy⁵⁰, Vincent Timmerman^{52,53}, Jonathan Baets^{54,55,56}, Liedewei Van de Vondel^{53,54}, Danique Beijer^{53,54}, Peter de Jonghe^{53,55}, Vincenzo Nigro^{57,58}, Sandro Banfi^{57,58}, Annalaura Torella⁵⁷, Francesco Musacchia^{57,58}, Giulio Piluso⁵⁷, Alessandra Ferlini⁵⁹, Rita Selvatici⁵⁹, Rachele Rossi⁵⁹, Marcella Neri⁵⁹, Stefan Aretz^{60,61}, Isabel Spier^{60,61}, Anna Katharina Sommer⁶⁰, Sophia Peters⁶⁰, Carla Oliveira^{62,63,64}, Jose Garcia Pelaez^{62,63}, Ana Rita Matos^{62,63}, Celina São José^{62,63}, Marta Ferreira^{62,63}, Irene Gullo^{62,63,64}, Susana Fernandes^{62,65}, Luzia Garrido⁶⁶, Pedro Ferreira^{62,63,67}, Fátima Carneiro^{62,63,64}, Morris A. Swertz¹⁸, Lennart Johansson¹⁸, Joeri K. van der Velde¹⁸, Gerben van der Vries¹⁸, Pieter B. Neerinx¹⁸, Dieuwke Roelofs-Prins¹⁸, Sebastian Köhler⁶⁸, Alison Metcalfe^{46,69}, Alain Verloes^{70,71}, Séverine Drunat^{70,71}, Caroline Roerck⁷², Aurelien Trimouille⁷³, Raffaele Castello⁵⁸, Manuela Morleo⁵⁸, Michele Pinelli⁵⁸, Alessandra Varavallo⁵⁸, Manuel Posada De la Paz⁷⁴, Eva Bermejo Sánchez⁷⁴, Estrella López Martín⁷⁴, Beatriz Martínez Delgado⁷⁴, F. Javier Alonso García de la Rosa⁷⁴, Andrea Ciolli⁷⁵, Bruno Dalla-piccola⁷⁵, Simone Pizzi⁷⁵, Francesca Clementina Radio⁷⁵, Marco Tartaglia⁷⁵, Alessandra Renieri^{76,77,78}, Elisa Benetti⁷⁶, Peter Balicza⁷⁹, Maria Judit Molnar⁷⁹, Ales Maver⁸⁰, Borut Peterlin⁸⁰, Alexander Münchau⁸¹, Katja Lohmann⁸¹, Rebecca Herzog⁸¹, Martje Pauly⁸¹, Alfons Macaya⁸², Anna Marcé-Grau⁸², Andres Nascimiento Osorio⁸³, Daniel Natera de Benito⁸³, Hanns Lochmüller^{84,85,86}, Rachel Thompson^{85,86}, Kiran Polavarapu⁸⁴, David Beeson⁸⁷, Judith Cossins⁸⁷, Pedro M. Rodriguez Cruz⁸⁷, Peter Hackman⁸⁸, Mridul Johari⁸⁸, Marco Savarese⁸⁸, Bjame Udd^{88,89,90}, Rita Horvath⁹¹, Gabriel Capella⁹², Laura Valle⁹², Elke Holinski-Feder⁹³, Andreas Laner⁹³, Verena Steinke-Lange⁹³, Evelin Schröck⁹⁴, Andreas Rump^{94,95}

²³Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands; ²⁴Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; ²⁵Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University Hospitals NHS Foundation Trust, Health Innovation Manchester, Manchester, UK; ²⁶Dijon University Hospital, Genetics Department, Dijon, France; ²⁷Dijon University Hospital, Centre of Reference for Rare Diseases, Development Disorders and Malformation Syndromes, Dijon, France; ²⁸Dijon University Hospital, FHU-TRANSLAD, Dijon, France; ²⁹Dijon University Hospital, GIMI Institute, Dijon, France; ³⁰University of Burgundy-Franche Comté, Dijon Economics Laboratory, Dijon, France; ³¹University of Burgundy-Franche Comté, FHU-TRANSLAD, Dijon, France; ³²EURORDIS-Rare Diseases Europe, Sant Antoni Maria Claret 167 - 08025, Barcelona, Spain; ³³EURORDIS-Rare Diseases Europe, Plateforme Maladies Rares, Paris, France; ³⁴Institut National de la Santé et de la Recherche Médicale (INSERM) U1127, Paris, France; ³⁵Centre National de la Recherche Scientifique, Unité Mixte de Recherche (UMR) 7225, Paris, France; ³⁶Unité Mixte de Recherche en Santé 1127, Université Pierre et Marie Curie (Paris 06), Sorbonne Universités, Paris, France; ³⁷Institut du Cerveau -ICM, Paris, France; ³⁸Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres Research University, Paris, France; ³⁹Centre de Référence de Neurogénétique, Hôpital de la Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France; ⁴⁰Hôpital de la Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France; ⁴¹AP-HP, Centre de Référence de Pathologie Neuromusculaire Nord, Est, Ile-de-France, Institut de Myologie, G.H. Pitié-Salpêtrière, Paris, France; ⁴²Institut de Myologie, Equipe Bases de données, G.H. Pitié-Salpêtrière, Paris, France; ⁴³AP-HP, Unité Fonctionnelle de Cardiogénétique et Myogénétique Moléculaire et Cellulaire, G.H. Pitié-Salpêtrière, Paris, France; ⁴⁴Department of Biology and Medical Genetics, Charles University Prague-2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; ⁴⁵Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; ⁴⁶Florence Nightingale Faculty of Nursing and Midwifery, King's College, London, UK; ⁴⁷Genetic Counselling, Genomics England, Queen Mary University of London, Dawson Hall, London, UK; ⁴⁸MRC Centre for Neuromuscular Diseases and National Hospital for Neurology and Neurosurgery, UCL Queen Square Institute of Neurology, London, UK; ⁴⁹Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK; ⁵⁰Dubowitz Neuromuscular Centre, UCL Great Ormond Street Hospital, London, UK; ⁵¹NIHR Great Ormond Street Hospital Biomedical Research Centre, London, UK; ⁵²Peripheral Neuropathy Research Group, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium; ⁵³Institute Born Bunge, Antwerp, Belgium; ⁵⁴Peripheral Neuropathy Research Group, University of Antwerp, Antwerp, Belgium; ⁵⁵Neuromuscular Reference Centre, Department of Neurology, Antwerp University Hospital, Antwerpen, Belgium; ⁵⁶Laboratory of Neuromuscular Pathology, Institute Born-Bunge, University of Antwerp, Antwerpen, Belgium; ⁵⁷Dipartimento di Medicina di Precisione, Università degli Studi della Campania "Luigi Vanvitelli", Napoli, Italy; ⁵⁸Telethon Institute of Genetics and Medicine, Pozzuoli, Italy; ⁵⁹Unit of Medical Genetics, Department of Medical Sciences, University of Ferrara, Ferrara, Italy; ⁶⁰Institute of Human Genetics, University of Bonn, Bonn, Germany; ⁶¹Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany; ⁶²IS - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; ⁶³IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; ⁶⁴Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal; ⁶⁵Department of Genetics, Faculty of Medicine, University of Porto, Porto, Portugal; ⁶⁶CHUSJ, Centro Hospitalar e Universitário de São João, Porto, Portugal; ⁶⁷Faculty of Sciences,

University of Porto, Porto, Portugal; ⁶⁸NeuroCure Cluster of Excellence, Charité Universitätsklinikum, Charitéplatz 1, Berlin, Germany; ⁶⁹College of Health, Well-being and Life-Sciences, Sheffield Hallam University, Sheffield, UK; ⁷⁰Department of Genetics, Assistance Publique-Hôpitaux de Paris - Université de Paris, Robert DEBRE University Hospital, 48 bd SERURIER, Paris, France; ⁷¹INSERM UMR 1141 “NeuroDiderot”, Hôpital R DEBRE, Paris, France; ⁷²Univ. Bordeaux, MRGM INSERM U1211, CHU de Bordeaux, Service de Génétique Médicale, Bordeaux, France; ⁷³Laboratoire de Génétique Moléculaire, Service de Génétique Médicale, CHU Bordeaux – Hôpital Pellegrin, Place Amélie Raba Léon, Bordeaux Cedex, France; ⁷⁴Institute of Rare Diseases Research, Spanish Undiagnosed Rare Diseases Cases Program (SpainUDP) & Undiagnosed Diseases Network International (UDNI), Instituto de Salud Carlos III, Madrid, Spain; ⁷⁵Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy; ⁷⁶Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Siena, Italy; ⁷⁷Medical Genetics, University of Siena, Siena, Italy; ⁷⁸Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy; ⁷⁹Institute of Genomic Medicine and Rare Diseases, Semmelweis University, Budapest, Hungary; ⁸⁰Clinical institute of genomic medicine, University medical centre Ljubljana, Ljubljana, Slovenia; ⁸¹Institute of Neurogenetics, University of Lübeck, Lübeck, Germany; ⁸²Neurology Research Group, Vall d’Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain; ⁸³Neuromuscular Disorders Unit, Department of Pediatric Neurology, Hospital Sant Joan de Déu, Barcelona, Spain; ⁸⁴Department of Neuropediatrics and Muscle Disorders, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany; ⁸⁵Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain; ⁸⁶Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada; ⁸⁷Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK; ⁸⁸Folkhälsan Research Centre and Medicum, University of Helsinki, Helsinki, Finland; ⁸⁹Tampere Neuromuscular Center, Tampere, Finland; ⁹⁰Vasa Central Hospital, Vaasa, Finland; ⁹¹Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK; ⁹²Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain; ⁹³Medical Genetics Center (MGZ), Munich, Germany; ⁹⁴Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, Technical University Dresden, Dresden, Germany; ⁹⁵Center for Personalized Oncology, University Hospital Carl Gustav Carus, Technical University Dresden, Dresden, Germany

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless

indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hartley T, Lemire G, Kernohan KD, Howley HE, Adams DR, Boycott KM. New diagnostic approaches for undiagnosed rare genetic diseases. *Annu Rev Genomics Hum Genet.* 2020;21: 351–72.
- EURORDIS AKFF. The Voice of 12,000 Patients. Experiences and Expectations of Rare Disease Patients on Diagnosis and Care in Europe. Eurordis; Paris, France; 2009.
- Smith HS, Swint JM, Lalani SR, Yamal JM, de Oliveira Otto MC, Castellanos S, et al. Clinical application of genome and exome sequencing as a diagnostic tool for pediatric patients: a scoping review of the literature. *Genet Med.* 2019;21:3–16.
- Wise AL, Manolio TA, Mensah GA, Peterson JF, Roden DM, Tamburro C, et al. Genomic medicine for undiagnosed diseases. *Lancet.* 2019;394:533–40.
- Liu P, Meng L, Normand EA, Xia F, Song X, Ghazi A, et al. Reanalysis of clinical exome sequencing data. *N. Engl J Med.* 2019;380:2478–80.
- Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, et al. Integrating healthcare and research genetic data empowers the discovery of 28 novel developmental disorders. *bioRxiv.* 2020. <https://doi.org/10.1101/797787>.
- Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature.* 2018;555:611–6.
- Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun.* 2017;8:15824.
- van der Velde KJ, Imhann F, Charbon B, Pang C, van Enckevort D, Slofstra M, et al. MOLGENIS research: advanced bioinformatics data software for non-bioinformaticians. *Bioinformatics.* 2019;35:1076–8.
- Zhao M, Havrilla JM, Fang L, Chen Y, Peng J, Liu C, et al. Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genom Bioinform.* 2020;2:lqaa032.
- Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacón A, et al. From wet-lab to variations: concordance and speed of bioinformatics pipelines for whole genome and whole exome sequencing. *Hum Mutat.* 2016;37:1263–71.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
- Matalonga L, Hernández-Ferrer C, Piscia D, Solve-RD SNV-indel working group, Vissers LELM, Schüle R, et al. Diagnosis of rare disease patients through programmatic reanalysis of genome-phenome data. Manuscript submitted to EJHG (703-20-EJHG).
- Boycott KM, Campeau PM, Howley HE, Pavlidis P, Rogic S, Oriel C, et al. The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. *Am J Hum Genet.* 2020;106:143–52.
- de Boer E, Ockeloen CW, Matalonga L, Horvath R, Solve-RD SNV-indel working group, Rodenburg RJ, et al. A pathogenic MT-TL1 variant identified by whole exome sequencing in an individual with unexplained intellectual disability, epilepsy and spastic tetraparesis. Manuscript submitted to EJHG (699-20-EJHG).

16. Schüle R, Timmann D, Erasmus CE, Reichbauer J, Wayand M, van de Warrenburg BPC, et al. Common pitfalls in genetic diagnosis of rare neurological diseases. Manuscript submitted to EJHG (705-20-EJHG).
17. Töpf A, Pyle A, Griffin H, Matalonga L, Schon K, Solve RD SNV indel working group, et al. Exome reanalysis and proteomic profiling identified TRIP4 as a novel cause of cerebellar hypoplasia and spinal muscular atrophy (PCH1). Manuscript submitted to EJHG (700-20-EJHG).
18. te Paske I, Garcia-Pelaez J, Sommer AK, Matalonga L, Starzynska T, Jakubowska A, et al. A Mosaic PIK3CA Variant in a Young Adult with Diffuse Gastric Cancer: Case Report. Manuscript submitted to EJHG (704-20-EJHG).
19. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.

Affiliations

Birte Zurek¹ · Kornelia Ellwanger¹ · Lisenka E. L. M. Vissers^{2,3} · Rebecca Schüle^{4,5} · Matthias Synofzik^{4,5} · Ana Töpf⁶ · Richarda M. de Voer^{2,7} · Steven Laurie⁸ · Leslie Matalonga⁸ · Christian Gilissen^{2,7} · Stephan Ossowski¹ · Peter A. C. 't Hoen^{7,9} · Antonio Vitobello¹⁰ · Julia M. Schulze-Hentrich¹ · Olaf Riess^{1,11} · Han G. Brunner^{2,3,12} · Anthony J. Brookes¹³ · Ana Rath¹⁴ · Gisèle Bonne¹⁵ · Gulcin Gumus¹⁶ · Alain Verloes¹⁷ · Nicoline Hoogerbrugge^{2,7} · Teresinha Evangelista¹⁵ · Tina Harmuth¹ · Morris Swertz¹⁸ · Dylan Spalding¹⁹ · Alexander Hoischen^{2,7,20} · Sergi Beltran^{8,21,22} · Holm Graessner^{1,11} · Solve-RD consortium

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

² Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands

³ Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

⁴ Department of Neurodegeneration, Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany

⁵ German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

⁶ John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research Institute, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

⁷ Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

⁸ CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

⁹ Center for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands

¹⁰ Inserm—University of Burgundy-Franche Comté, Dijon, France

¹¹ Centre for Rare Diseases, University of Tübingen, Tübingen, Germany

¹² Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands

¹³ Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

¹⁴ INSERM, US14—Orphanet, Plateforme Maladies Rares, Paris, France

¹⁵ Sorbonne Université, INSERM UMRS 974, Center of Research in Myology, Paris, France

¹⁶ EURORDIS-Rare Diseases Europe, Barcelona, Spain

¹⁷ Genetics Department, APHP-Robert Debré University Hospital, Université de Paris, Paris, France

¹⁸ Department of Genetics, Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

¹⁹ European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge, UK

²⁰ Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands

²¹ Universitat Pompeu Fabra (UPF), Barcelona, Spain

²² Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain