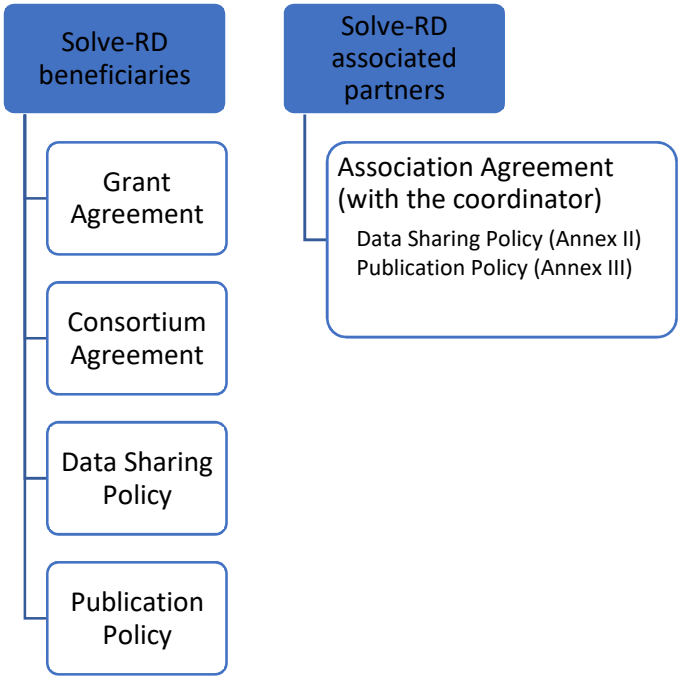
Solve-RD Code of Conduct

Version	V2.1
Date	18.12.2020 (V2.1) 19.06.2020 (V2) 23.01.2020 (V1)
History of changes	V2.1: Update Attachment 3 (Solve-RD Publication Policy) V2: Footnote added on page 3 (download from sftp server)
Authors	Holm Graessner (EKUT), Birte Zurek (EKUT), Ana Topf (UNEW), Sergi Beltran (CNAG-CRG), Lennart Johansson (UMCG), Martina Melovic (EKUT, legal council)
Approved by GA	22.01.2020 (V1)

Definitions

- Investigator: Individual data user affiliated to a Solve-RD beneficiary or associated partner.
- Raw dataset: exome or genome sequencing data, omics data, phenotype data from single individuals which have not been processed through an analysis pipeline.
- Processed dataset: data that are the result of raw data being processed through an analysis pipeline by Solve-RD or Solve-RD partners. It also includes analysis results.
- Solve-RD Sandbox: temporary secure compute environment / private 'cloud' for storage of data as well as tailored and *de novo* bioinformatics analysis in Solve-RD. The Solve-RD sandbox is composed of two parts: 1) an EGA box where raw data is being stored and 2) an European Bioinformatics Institute hosted analysis sandbox which can be used for data analysis. All data in the analysis sandbox (part 2 above) will be deleted on the completion of the Solve-RD project unless Union or Member State law requires storage of the data. Data in the EGA box (part 1 above) will be properly archived for the future at the EGA to enable controlled access distribution.

Contractual basis



The following is the Code of Conduct that the Investigator agrees to abide by as a user of the Solve-RD analysis sandbox. Failure to abide by any term within this Code of Conduct may result in revocation of approved access to any or all data obtained through the Sandbox.

Solve-RD comprises pseudonymised sensitive data of rare disease patients and their family members (phenotype data, metadata, raw datasets, and processed datasets) that are provided by contributing beneficiaries and associated partners. The contributing beneficiaries and associated partners are the primary data controllers of the datasets they contribute. Data shall not be processed, downloaded and/or used for any analysis unless the specific analysis/usage has been registered through the Solve-RD *Analysis Template* (see attachment 1) and the Investigator has been granted user rights for the Sandbox. Specifically, Investigators will work according to the following rules:

Principles

1. Investigator will adhere to the Solve-RD Data Sharing Policy[1] (see attachment 2).
2. Investigator will adhere to the Solve-RD Publication Policy[2] (see attachment 3).
3. Investigator acknowledges Solve-RD funding in any publication stemming from data processing of any Solve-RD data using the following wording: "The Solve-RD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257.".

Access to the Sandbox

4. Investigator's permission to access Solve-RD data is limited to all Investigators that have signed this Code of Conduct.

[1] Latest version available here: http://solve-rd.eu/results/presentations/
[2] Latest version available here: http://solve-rd.eu/results/presentations/

5. Investigator will refrain from downloading data to any location that has not been approved by Solve-RD[3].
6. Approvals for downloading data from the Solve-RD Sandbox to secure local clusters will be granted by the Solve-RD Steering Committee upon application from Investigator. Record of all persons locally accessing the data must be kept by the Investigator who has been granted approval and shared with Solve-RD upon request.
7. Downloaded data will have to be deleted on the completion of the Solve-RD project unless Union or Member State law requires storage of the data. Investigator will confirm the deletion of downloaded data and respective processed data to the Solve-RD project office.
8. Investigator will report any inadvertent data release, breach of data security, or other data management incidents contrary to the terms of data access specified in this Code of Conduct to info@solve-rd.eu.

Data processing

9. Investigator will perform all analyses solely on the Sandbox, on approved local clusters ensuring data security or on a Solve-RD approved secure space.
10. Investigator will process datasets contained in the Solve-RD Sandbox solely in connection with an approved analysis project described in the Solve-RD Analysis Template (see attachment 1).
11. When data is downloaded from the Sandbox to a local cluster for processing locally, Investigator will ensure that all downloaded data as well as intermediate/output data generated through analysis/data usage, are solely accessible by the Investigator him or herself and their team.
12. Investigator agrees that the data may only be used for the permitted purpose to conduct the approved analysis/data usage.
13. Investigator is not allowed to inform the primary controller of the dataset about an incidental finding, unless the primary controller of that dataset explicitly requests this information from Investigator.
14. Investigator will provide the results of the analysis/data usage to the primary controller of the dataset as soon as the analysis results are available.
15. Investigator is aware that all activities with regards to processing data in the Sandbox and downloading data will be monitored at the level of individual researcher.
16. Investigator will make no attempt to identify or contact participants from whom these data were collected unless re-contact is appropriate for particular reasons. Such decision will be made by the Solve-RD SC. Participants will have the right to withdraw.

I read and understand the Solve-RD Code of Conduct as well as the Solve-RD Data Sharing Policy and Publication Policy. I declare that I will abide by it and will act accordingly. I understand that failure to adhere to the Code of Conduct will have repercussions for my role in Solve-RD.

Please, sign this Code of Conduct and send a scanned copy to birte.zurek@med.uni-tuebingen.de

[3] Download of analysis result files from the DITF folders on the sftp server by DITF and WG members is allowed.

Attachment 1: Data analysis template (Version 2 | 12.03.2020)

Project Template for Usage and/or Analysis of data within Solve-RD[4]

Investigators	*Person responsible for the analysis project (WG lead)* *Other key personnel / external collaborators involved (WG members)*	
Affiliated institutions	*List all institutions involved*	
Email address of responsible investigator	*Address of responsible person*	
Date	*dd.mm.yyyy (date of application)*	
Title of the proposed project		
Project Title		
Brief description of the project		
Keywords	*Add max 5 keywords.*	
Specific aims	*State the aims of the project as specifically as possible.*	
Study design	*Describe the study design.*	
Cohort, clinical / OMICS data	*Summarise the data you will use/analyse.* *Describe inclusion criteria for cohort/data selection and approximate numbers if applicable.*	
Plans for analysis / use		
Required analysis resources	*Summarise the storage and/or compute capacity needed for the analysis.* *Clarify if you will do analyses in the sandbox and/or local cluster.*	
Download of data	Date of download:	*If applicable [dd.mm.yyyy]*
	Location of the secure local cluster where data is being stored:	*If applicable*
	Downloaded datasets:	*If applicable*
End date	*Provide estimated date [dd.mm.yyyy] by when this project will be completed.*	

[4] Please do only fill out the white fields.

Contributions	*Please list main contributors for this project (DITFs, specific groups etc.).*

Please fill in this template and send it to the DITF coordinators of all ERNs concerned and the Solve-RD coordination office (birte.zurek@med.uni-tuebingen.de). DITF coordinators will inform the respective data submitters of their DITF. If DITF coordinators deem the planned analysis as being not in accordance with the Solve-RD objectives and Data Sharing Policy, they will express their concerns to the WP2 leads Alex Hoischen and Sergi Beltran.

Attachment 2: Solve-RD Data Sharing Policy

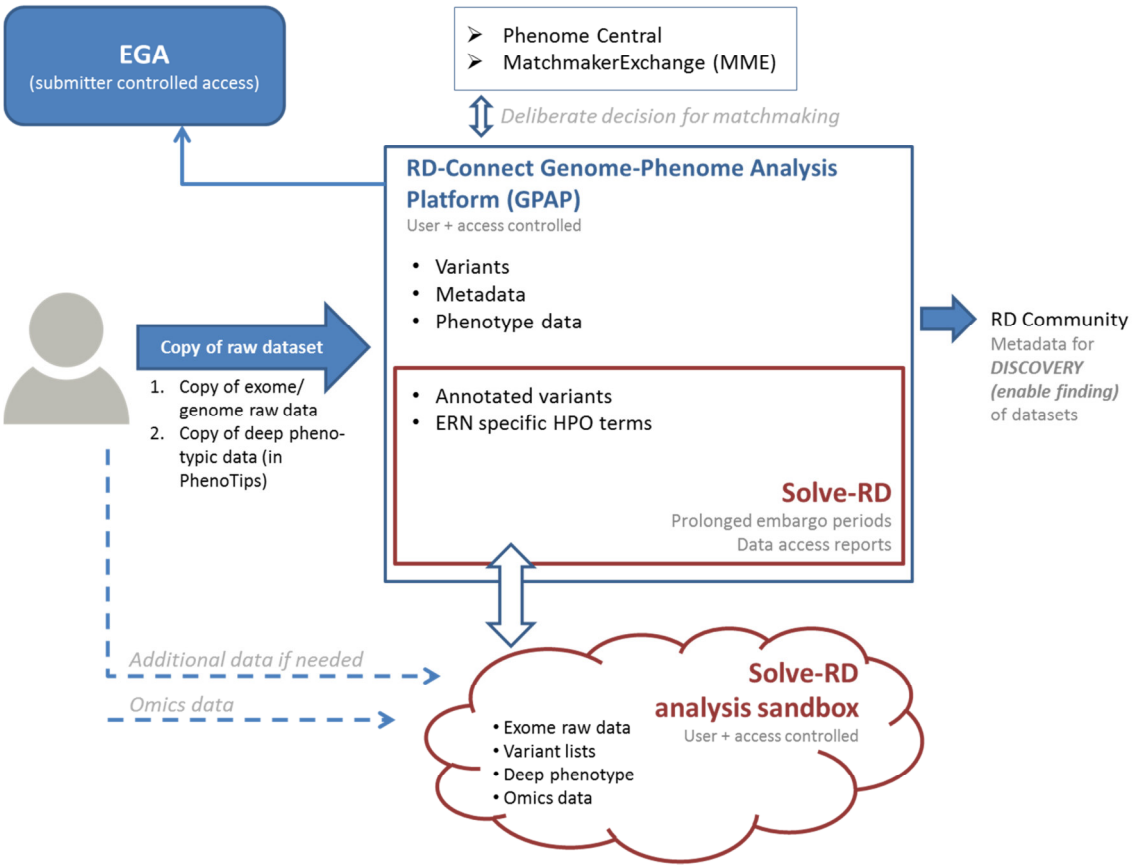
Solve-RD Data Sharing Policy

Version	Final V7
Date	09.01.2020
Authors	EKUT: Holm Graessner, Tina Harmuth, Birte Zurek, Martina Melovic CNAG-CRG: Sergi Beltran ULEIC:Tony Brookes, UMCG: Morris Swertz
Approved by GA	22.01.2020

Definitions

- Raw dataset: exome or genome sequencing data in FASTQ or BAM format with corresponding phenotype data in HPO format from a single individual.
- Data collection: multiple raw datasets from a single centre.
- Processed dataset: a raw dataset whose exome or genome sequencing data has been processed through the Solve-RD/RD-Connect variant calling analysis pipeline. It includes the variant calls, some metrics and the corresponding phenotypic information.
- RD-Connect Genome-Phenome Analysis Platform (GPAP): platform that enables collation, processing, analysis, interpretation and sharing of integrated genome and phenome datasets. The RD-Connect GPAP will only retain processed datasets (https://platform.rd-connect.eu/).
- European Genome-phenome Archive (EGA): storage facility that is used by the RD-Connect GPAP for storage of the raw datasets (https://ega-archive.eu).
- Solve-RD analysis sandbox: secure compute environment / private 'cloud' for tailored and *de novo* bioinformatics analysis. Data will be deleted after the Solve-RD analysis have been finalized that is at the latest two years after the end of the project (i.e. after December 31st 2024).
- PhenoTips: user friendly tool integrated in the RD-Connect GPAP to collate and store the phenotypic information from a dataset using ontologies and standards such as HPO, ORDO and OMIM (https://platform.rd-connect.eu/, https://phenotips.org).

Explanation figure of data sharing process



Most important points of Solve-RD data sharing policy

- All exome/genome and phenotypic data that will be collated in task 1.2 of Solve-RD will be submitted to the RD-Connect Genome-Phenome Analysis Platform (GPAP)
- All IT systems used in Solve-RD comply with the General Data Protection Regulation, GDPR (Regulation (EU) 2016/679)
- Options for uploading datasets include single dataset upload and bulk upload
- Specific data access stipulations for Solve-RD in GPAP:
  - All data submitters will be able to see which other users have accessed their submitted datasets and when
  - If justified Solve-RD data submitters can define longer embargo periods before data become accessible to other users

Upload of exome/genome and phenotypic data to the RD-Connect GPAP

All exome/genome and phenotypic data that will be collated in task 1.2 of Solve-RD will be submitted to the RD-Connect Genome-Phenome Analysis Platform, an IRDiRC recognised resource. This empowers the clinical end-users and their research teams to analyse and interpret their own data and actively participate in solving cases rather than only handing it over to the project and waiting for results.

Where unsolved processed datasets from previous projects are already available in the platform, there will be an option to assign these to the Solve-RD project, while also keeping the assignment to the original project (if this is the case). An option to assign a new or existing dataset to a specific European Reference Network (ERN) will also be available.

Options for uploading

One raw dataset is considered to include both **phenotypic data** in the form of HPO terms and **exome/genome data** of a patient, preferentially in FASTQ format, although BAM files are also accepted.

There are two options for uploading datasets.

i. **Data collections from centres that will upload fewer than 100 raw datasets**: raw datasets will be uploaded using the standard RD-Connect GPAP upload interface, which includes user friendly PhenoTips templates to enter the phenotypic information and user-friendly tables to upload the genomic data and metadata. An option to bulk upload the genomic data and metadata is also available.

ii. **Data collection from centres that will upload 100 or more raw datasets**: in cooperation with CNAG-CRG in Barcelona there will be the opportunity to discuss a customised bulk upload option to facilitate uploading large data collections.

All sequencing data are submitted as raw data in FASTQ (or BAM) format and are processed through the same Solve-RD pipeline. The processed datasets will be stored in the platform for as long as it has a legitimate interest in doing so. (see below for storage of raw data at the EGA). Clinical interpretation of the data (final or temporary) can be entered in the RD-Connect platform but this is not a requirement at upload.

The raw data will not be stored in the European Genome-phenome Archive (see below) for longer than is necessary. Personal data will be processed fairly and lawfully in accordance with the EU General Data Protection Regulation 2016/679.

Data security and accessibility

RD-Connect pays strict attention to data quality and security and the data in the Platform meet high quality and safety standards. The RD-Connect registration process includes user validation as defined in the RD-Connect GPAP Code of Conduct (https://rd-connect.eu/gpap-code-conduct), which all users must confirm they accept. Additionally, the PI/group leads must sign the Adherence Agreement.

Other computer systems used in Solve-RD will adhere to similar standards.

(i) Who can access the processed dataset/s in the RD-Connect Genome-Phenome Analysis Platform?

- Data submitters can request an embargo period for each of their datasets. During the embargo period the data is accessible only to the members from the submitter group and the Solve-RD user group. However, the members of the original submitter group can share it specifically with other group/s of users (from another PI/group lead).
- After the embargo period set by the submitter, datasets uploaded to the RD-Connect GPAP will become accessible to all authorised scientists and clinicians who have gone through the strict registration and verification process.
- The embargo period is considered to start at the moment a specific processed dataset (genomic plus phenotypic data) is made accessible to the dataset submitter.

(ii) Who can access Solve-RD processed datasets within the RD-Connect GPAP and under which conditions?

- Account registration: Every PI/group lead enrolled in Solve-RD or in a participating ERN will undergo the full RD-Connect registration process and will then enrol members of their team. All the users under the responsibility of one PI/group lead are assigned to the same user group and have the same user permissions but have different usernames and passwords to enable user-specific logs. In addition, every PI/group that will upload samples for Solve-RD and that is not a Solve-RD beneficiary will have to sign an association agreement with Solve-RD containing the Solve-RD Data Sharing Policy and Publication Policy.
- Solve-RD tagging of datasets: New datasets uploaded specifically for the Solve-RD project will be assigned to the Solve-RD project to allow project-wide sharing and monitoring. Pre-existing unsolved processed datasets can be added to the Solve-RD project (while also keeping the original project tag). Datasets must also be tagged with the name of the submitting European Reference Network (ERN) so that it is possible to follow numbers submitted per ERN.
- No embargo period within Solve-RD: Data become accessible to all other authorised Solve-RD users immediately after submission. Solve-RD users can analyse and query their own datasets as well as datasets submitted by other Solve-RD users.
- Access of Solve-RD datasets by other users of the RD-Connect Genome-Phenome Analysis Platform that are not part of Solve-RD:
  i. Embargo periods of up to twelve months: As part of the online data submission process, Solve-RD users can easily define an embargo period of up to twelve months before data become accessible to other users of the RD-Connect Genome-Phenome Analysis Platform. Solve-RD users selecting and embargo period of up to six months will not be asked for any additional information, but Solve-RD data submitters requesting an embargo period between six and twelve months will be asked to type a short justification. RD-Connect has agreed to delegate approval of embargo periods between six and twelve months to the Solve-RD Steering Committee, which will only contact Solve-RD data submitters if the embargo period has been denied or more details are needed.
  ii. Embargo periods longer than twelve months: depending on the origin and nature of data (for example if they are diagnostic data) Solve-RD users can request (in written form at the time of submission) an embargo period longer than twelve months before data become accessible to other users of the RD-Connect Genome-Phenome Analysis Platform. Requests longer than twelve months will require acceptance by the Solve-RD Steering Committee and by the RD-Connect GPAP Access Committee since they are considered non-compliant with IRDiRC principles for rapid data release. Retrospective requests for prolonged embargo periods are not possible.
- Access reports: within the RD-Connect GPAP data management portal, a dataset submitter is able to see which other users have accessed at which date a given dataset through a specific query on that dataset or through a general query.

(iii) What can users do with the data?

- During embargo: Datasets are only accessible to the submitter and the Solve-RD project members. Members from the submitter group can share specific datasets with other RD-Connect GPAP users. Users with access to the datasets will be able to discover, query, analyse, interpret and tag them. If the submitter group has opened the dataset to matchmaking through MatchMaker Exchange, internal RD-Connect users and external users across the globe

performing a matchmaking query may be informed that there is a dataset containing a potential match, but the user cannot see the relevant dataset and must contact the submitter to find out more details or request sharing.

- After embargo: Datasets are accessible to the other authorised users within the RD-Connect GPAP, who will also be able to discover, query, analyse, interpret and tag them. The datasets also need to be specifically opened by the submitter group for matchmaking.
- Download of data: Direct download of full datasets is not possible at any time. Download of search results will be restricted to the user group that submitted the dataset and users with whom it may have been specifically shared by the submitter group.

(iv) Data security

Data is stored in a computer cluster with a restricted access policy, limited internet access and daily backups. Databases are using distributed filesystems, limiting the risk of physical attacks. All communications are encrypted. Security of the platform was audited in October 2017 with no major risks being identified. Platform requests and user actions are safely logged for audit purposes. Documentation and procedures are currently being adapted for the new General Data Protection Regulation, GDPR (Regulation (EU) 2016/679).

Matchmaking and discovery

The RD-Connect GPAP is integrated in the Beacon Network (https://beacon-network.org), a project by the Global Alliance for Genomics and Health (GA4GH).

RD-Connect participates in MatchMaker Exchange (MME, http://www.matchmakerexchange.org) and MME is functional for internal GPAP queries and bi-lateral queries to PhenomeCentral. Bi-lateral queries to DECIPHER are being implemented. Additionally, patient profiles can be pushed to PhenomeCentral (https://www.phenomecentral.org) from the RD-Connect GPAP PhenoTips instance by the data submitters. The dataset submitter must specifically enable matchmaking at the time of submission or in the data management portal. This permission can also be enabled or disabled at any later stage. Increasing powerful and precise forms of matchmaking will be developed by Solve-RD in conjunction with GA4GH and others, to enable ever more sophisticated dataset discovery and matchmaking with more options for data protection. Dataset submitters will have to specifically enable the datasets for new forms of discovery and/or matchmaking outside the RD-Connect GPAP and/or Solve-RD if they are less restrictive than the current MME v1.

Patient security and confidentiality

To protect patient privacy, explicitly identifiable patient information is never submitted to the Platform. The submitting clinician stores the data in the manner appropriate for their own centre and links it at a secure local level to the unique RD-Connect IDs. Patient identities are therefore not stored on the Platform and cannot be accessed by Platform users. Only the researcher who submitted the data has the key to identify the IDs corresponding to his/her data.

(v) Return of results

Solve-RD is a research project and as such cannot ensure the quality standards required for genetic diagnostics. It is thus the responsibility of the clinician/researcher who submitted the data and who is the only one who has access to the patient and family to validate any novel genes and to return the results to the patient and her/his family (depending on the consent given).

It is possible that other GPAP users identify genetic variants in sequencing data submitted by Solve-RD partners. These variants may explain the cause of the patients' disease but they may also be completely unrelated – so called 'incidental findings'. Some clinicians don't want to be informed about such incidental

findings. We have thus put a system in place where the data submitter can indicate for every patient if they DO NOT WANT to be notified about (forced) incidental findings: "This patient has not consented to be notified about incidental findings and I don't want to be contacted regarding incidental findings on this patient". The system to contact other users, which would basically send an email to both data submitter and the "data analyst" to put them in contact on a certain experiment, includes this message so that the "data analyst" would know that he/she should NOT contact the submitter for any findings which are not related to the patients' disease.

(vi) Storage of raw data in the European Genome-phenome Archive (EGA)

The raw data will not be kept for longer than is necessary for the purposes at the European Genome-phenome Archive (EGA), a secure, controlled-access repository. The EGA serves as an archive for publications as well as data on several levels, including the raw data (to allow future reanalysis using other algorithms) and the genotype calls (information about pathogenic genetic variants) provided by the data submitters.

The EGA provides the necessary security required to control access to the data and maintain patient confidentiality. Data can be accessed only by authorised researchers and clinicians. In all cases, data access decisions are made not by the EGA but by an appropriate Data Access Committee, which can be the person or group submitting the data.

Data must be submitted to the EGA at the latest by the end of Solve-RD. At the time of uploading a dataset to the RD-Connect GPAP, the user can indicate if the dataset is already available at the EGA and provide the corresponding reference number. For datasets not yet available at the EGA, the CNAG-CRG will broker the submission to the EGA of the data and metadata uploaded to the RD-Connect GPAP. The original data submitter will be responsible for making decisions regarding the future access to their datasets.

(vii) Solve-RD publications - notification and authorship policy with regard to shared data

All Solve-RD publications are acknowledged to be based on the fundamental principles of open scientific collaboration, reciprocity, attribution and benefit sharing. For any publication resulting from work carried out using data shared or generated through Solve-RD (e.g. for identifying a novel gene), including where data has been accessed through the RD-Connect Genome-Phenome Analysis Platform, the Analysis Sandbox or future Solve-RD systems, the authors should in all cases acknowledge and give appropriate authorship positions to all relevant parties in line with best practice for acknowledgement of scientific contribution including submission of the primary data (also see the Solve-RD Publication Policy).

Examples and further principles are described below.

1. **A publication arising from research in which the party leading the publication ("the PI team") is primarily analysing their own submitted data (example: novel gene discovery by a submitter analysing their own patient cohorts in the RD-Connect GPAP):**

   i. Where a publication only includes data and hypotheses from the PI's own research group, key authorship positions may be held by this group, but the software, tools and resources made use of for the research should be duly acknowledged and referenced in line with the policies for those resources (e.g. see RD-Connect GPAP policy below). Where justified, individuals supporting the bioinformatics analysis or platforms may be approached for co-authorship based on individual scientific contribution.

   ii. Where a publication has involved the use or analysis of data from additional submitters, these submitters should be contacted as soon as possible ahead of publication and invited to provide input as co-authors. The PI team is strongly encouraged to share key authorship positions with other teams that have brought in similar intellectual input and/or fundamental data (e.g. "a

second family"). Acknowledgement of bioinformatics support should also be considered as in (i) above.

2. **A publication arising from the analysis of data where the party generating the hypothesis and carrying out the analysis is not themselves the data submitter (example: reanalysis of data by a Solve-RD bioinformatics group that did not submit the data or see the patients):**

   i. Submitters of the data used for the analysis should be contacted as soon as possible ahead of publication and invited to provide input as co-authors. The PI team is strongly encouraged to share key authorship positions with the submitting teams based on the value and amount of data contributed to the publication. If the primary data is the key to discovery, a key authorship position should be discussed with the owner of the primary data.

   ii. Where a publication makes use of data from a large number of submitters or transversal analysis of the Solve-RD cohort, a group authorship for Solve-RD (see Solve-RD Publication Policy) should be considered in order to acknowledge the role of all data submitters equally.

All data access through the RD-Connect Genome-Phenome Analysis Platform is monitored automatically by the system and all other data access for other Solve-RD activities is only to named individuals within the Solve-RD consortium, therefore any breach of the publication policy will be monitored and flagged up to the Solve-RD Steering Committee.

Solve-RD funding acknowledgement

Any publications arising from Solve-RD project funding should acknowledge it in the following way:

> "This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257 (Solve-RD)."

RD-Connect GPAP acknowledgement

In addition to authorship positions as described above, any publications that arise from the use of the RD-Connect Genome Phenome Analysis Platform should acknowledge it in the following way:

> "This study makes use of data shared/provided through RD-Connect, which received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444."

In addition, the following paper should be cited:

> Lochmüller H & Badowska D, Thompson R, Knoers N, Aartsma-Rus A, Gut I, Wood L, Harmuth T, Durudas A, Graessner H, Schaefer F & Rieß O. RD-Connect, NeurOmics and EURenOmics: Collaborative European Initiative for Rare Diseases. European Journal of Human Genetics. 2018.

ERN acknowledgement

Any publications with contributions from ERNs should acknowledge involved ERNs in the following way:

> "This study was supported by the European Reference Network(s) [add ERN names] (https://ec.europa.eu/health/ern/networks_en)."

Attachment 3: Solve-RD Publication Policy

Solve-RD Publication Policy

Version	V2.1
Date	18 December 2020 (V2.1) 18 April 2018 (V2)
History of changes	V2.1: Update Annex I (Solve-RD author list)
Authors	Holm Graessner, Tina Harmuth
Approved by SC	24.4.2018 (V2)

1 Solve-RD as an author (if allowed for by journal)

An affiliation list will be created and regularly updated and will contain all members of the Solve-RD consortium (see Annex I). This list will determine the Solve-RD author ('Solve-RD consortium'). The author list will be reviewed and updated with each publication submission that makes use of the author list by the project management team (lead: Holm Graessner). The Solve-RD author shall be used for all publications if allowed for by the journal.

2 Notification and authorship policy with regard to shared data

All Solve-RD publications are acknowledged to be based on the fundamental principles of open scientific collaboration, reciprocity, attribution and benefit sharing. For any publication resulting from work carried out using data shared or generated through Solve-RD (e.g. for identifying a novel gene), including where data has been accessed through the RD-Connect Genome-Phenome Analysis Platform, the authors should in all cases acknowledge and give appropriate authorship positions to all relevant parties in line with best practice for acknowledgement of scientific contribution including submission of the primary data.

Examples and further principles are described below.

3. **A publication arising from research in which the party leading the publication ("the PI team") is primarily analysing their own submitted data (example: novel gene discovery by a submitter analysing their own patient cohorts in the RD-Connect GPAP):**

   iii. Where a publication only includes data and hypotheses from the PI's own research group, key authorship positions may be held by this group, but the software, tools and resources made use of for the research should be duly acknowledged and referenced in line with the policies for those resources (e.g. see RD-Connect GPAP policy below). Where justified, individuals supporting the bioinformatics analysis or platforms may be approached for co-authorship based on individual scientific contribution.

   iv. Where a publication has involved the use or analysis of data from additional submitters, these submitters should be contacted as soon as possible ahead of publication and invited to provide input as co-authors. The PI team is strongly encouraged to share key authorship positions with other teams that have brought in similar intellectual input and/or fundamental data (e.g. "a second family"). Acknowledgement of bioinformatics support should also be considered as in (a) above.

4. **A publication arising from the analysis of data where the party generating the hypothesis and carrying out the analysis is not themselves the data submitter (example: reanalysis of data by a Solve-RD bioinformatics group that did not submit the data or see the patients):**

   iii. Submitters of the data used for the analysis should be contacted as soon as possible ahead of publication and invited to provide input as co-authors. The PI team is strongly encouraged to share key authorship positions with the submitting teams based on the value and amount of

data contributed to the publication. If the primary data is the key to discovery, a key authorship position should be discussed with the owner of the primary data.

iv. Where a publication makes use of data from a large number of submitters or transversal analysis of the Solve-RD cohort, a group authorship for Solve-RD should be considered in order to acknowledge the role of all data submitters equally.

All data access through the RD-Connect genome-phenome analysis platform is monitored automatically by the system and all other data access for other Solve-RD activities is only to named individuals within the Solve-RD consortium, therefore any breach of the publication policy will be monitored and flagged up to the Solve-RD Steering Committee.

3 Confirmation of paper by Solve-RD Consortium

During the project and for a period of one (1) year after the project every paper that is published with affiliation of Solve-RD or includes data produced or collated within Solve-RD has to be confirmed by the Solve-RD Consortium. The procedure is defined in Article 29.1 of the Solve-RD Grant Agreement and Article 8.4.2 of the Solve-RD Consortium Agreement. All Parties and associated Partners (including associated ERNs) are obliged to follow this procedure.

Prior notice of any planned publication shall be given to the other Parties at least 45 calendar days before the intended date of publication. Any objection to the planned publication following the above notification shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within thirty (30) calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted.

An objection is justified if (a) the protection of the objecting Party's Results or Background would be adversely affected and/or (b) the objecting Party's legitimate academic or commercial interests in relation to the Results or Background would be significantly harmed.

The objection has to include a precise and reasonable request for necessary modifications, it being specified that any such modifications shall not harm the scientific content of the proposed publication or communication.

If an objection has been raised the involved Parties shall discuss how to overcome the justified grounds for the objection on a timely basis (for example by amendment to the planned publication and/or by protecting information before publication) and the objecting Party shall not unreasonably continue the opposition if appropriate measures are taken following the discussion. The objecting Party can request a publication delay of not more than 90 calendar days from date of submission to the other Parties. After 90 calendar days the publication is permitted provided that Confidential Information of the objecting Party has been removed from the Publication and all reasonable modifications of the objecting Party have been implemented within the Publication as indicated by the objecting Party.

A decision has to be made and communicated within four weeks after submission of the publication draft (author list and abstract) to the Steering Committee.

4 Parallel Analysis of submitted data at centres

The Steering Committee has to be informed if data sets that were submitted to Solve-RD for central analysis go through in house analysis that may lead to publications.

5 Acknowledgements

Solve-RD funding acknowledgement

Any publications arising from Solve-RD project funding should acknowledge it in the following way:

> "This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257 (Solve-RD)."

RD-Connect GPAP acknowledgement

In addition to authorship positions as described above, any publications that arise from the use of the RD-Connect Genome Phenome Analysis Platform should acknowledge it in the following way:

> "This study makes use of data shared/provided through RD-Connect, which received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444."

In addition, the following paper should be cited:

> Lochmüller H & Badowska D, Thompson R, Knoers N, Aartsma-Rus A, Gut I, Wood L, Harmuth T, Durudas A, Graessner H, Schaefer F & Rieß O. RD-Connect, NeurOmics and EURenOmics: Collaborative European Initiative for Rare Diseases. European Journal of Human Genetics. 2018.

ERN acknowledgement

Any publications with contributions from ERNs should acknowledge involved ERNs in the following way:

> "This study was supported by the European Reference Network(s) [add ERN names] (https://ec.europa.eu/health/ern/networks_en)."

Annex I: Solve-RD author list

EKUT: Olaf Riess[1,2], Tobias B. Haack[1] , Holm Graessner[1,2], Birte Zurek[1,2], Kornelia Ellwanger[1,2], Stephan Ossowski[1], German Demidov[1], Marc Sturm[1], Julia M. Schulze-Hentrich[1], Rebecca Schüle[3,4], Christoph Kessler[3,4] , Melanie Wayand[3,4], Matthis Synofzik[3,4], Carlo Wilke[3,4], Andreas Traschütz[3,4], Ludger Schöls[3,4], Holger Hengel[3,4], Peter Heutink[3,4].

RUMC: Han Brunner[5,6,7], Hans Scheffer[5,6], Nicoline Hoogerbrugge[5,8] , Alexander Hoischen[5,8,9], Peter A.C. 't Hoen[8,10] ,Lisenka E.L.M. Vissers[5,7], Christian Gilissen[5,8], Wouter Steyaert[5,8], Karolis Sablauskas[5], Richarda M. de Voer[5,8], Erik-Jan Kamsteeg[5], Bart van de Warrenburg[7,11], Nienke van Os[7,11] , Iris te Paske[5,8], Erik Janssen[5,8], Elke de Boer[5,7], Marloes Steehouwer[5], Burcu Yaldiz[5], Tjitske Kleefstra[5,7].

University of Leicester: Anthony J. Brookes[12], Colin Veal[12], Spencer Gibson[12], Marc Wadsley[12], Mehdi Mehtarizadeh[12], Umar Riaz[12], Greg Warren[12], Farid Yavari Dizjikan[12], Thomas Shorter[12].

UNEW: Ana Töpf[13], Volker Straub[13], Chiara Marini Bettolo[13], Sabine Specht[13].

MUH: Jill Clayton-Smith[14], Siddharth Banka[14,15], Elizabeth Alexander[14], Adam Jackson[14].

DIJON: Laurence Faivre[16,17,18,19,20] Christel Thauvin[17,18,19,20,20], Antonio Vitobello[18], Anne-Sophie Denommé-Pichon[18], Yannis Duffourd[18,19], Emilie Tisserant[18], Ange-Line Bruel[18], Christine Peyron[21,22], Aurore Pélissier[22,22].

CNAG-CRG: Sergi Beltran[23,24], Ivo Glynne Gut[24,24], Steven Laurie[24], Davide Piscia[24], Leslie Matalonga[24], Anastasios Papakonstantinou[24], Gemma Bullich[24], Alberto Corvo[24], Carles Garcia[24], Marcos Fernandez-Callejo[24], Carles Hernández[24], Daniel Picó[24], Ida Paramonov[24], Hanns Lochmüller[24].

EURORDIS: Gulcin Gumus[25], Virginie Bros-Facer[26].

INSERM-Orphanet: Ana Rath[27], Marc Hanauer[27], Annie Olry[27], David Lagorce[27], Svitlana Havrylenko[27], Katia Izem[27], Fanny Rigour[27].

INSERM-ICM: Giovanni Stevanin[28,29,30,31,32], Alexandra Durr[29,30,31, 31,33], Claire-Sophie Davoine[29,30,31,32,32], Léna Guillot-Noel[29,30,31,32,32], Anna Heinzmann[29,30,31,31,34], Giulia Coarelli[29,30,31,31,34].

INSERM-CRM: Gisèle Bonne[35], Teresinha Evangelista[35], Valérie Allamand[35], Isabelle Nelson[35], Rabah Ben Yaou[35,36,37], Corinne Metay[35,38], Bruno Eymard[35,36], Enzo Cohen[35], Antonio Atalaia[35], Tanya Stojkovic[35,36].

Univerzita Karlova: Milan Macek Jr.[39], Marek Turnovec[39], Dana Thomasová[39], Radka Pourová Kremliková[39], Vera Franková[39], Markéta Havlovicová[39], Vlastimil Kremlik[39].

EMBL-EBI: Helen Parkinson[40], Thomas Keane[40], Dylan Spalding[40], Alexander Senf[40].

Jackson Laboratory: Peter Robinson[41], Daniel Danis[41].

KCL: Glenn Robert[42], Alessia Costa[4242], Christine Patch[4242,43].

UCL-IoN: Mike Hanna[44], Henry Houlden[45], Mary Reilly[44], Jana Vandrovcova[45].

UCL-ICH: Francesco Muntoni[46,47], Irina Zaharieva[46], Anna Sarkozy[46].

Universiteit Antwerpen: Vincent Timmerman[48,49], Jonathan Baets[50,51,52], Liedewei Van de Vondel[49,50], Danique Beijer[49,50], Peter de Jonghe[49,51].

Uni Naples: Vincenzo Nigro[53,54], Sandro Banfi[53,54], Annalaura Torella[53], Francesco Musacchia[53,54], Giulio Piluso[53].

UNIFE: Alessandra Ferlini[55], Rita Selvatici[55], Rachele Rossi[55], Marcella Neri[55].

UKB: Stefan Aretz[56,57], Isabel Spier[56,57], Anna Katharina Sommer[56], Sophia Peters[56].

IPATIMUP: Carla Oliveira[58,59,60], Jose Garcia Pelaez[58,59], Ana Rita Matos[58,59], Celina São José[58,59], Marta Ferreira[58,59], Irene Gullo[58,59,60], Susana Fernandes[58,61], Luzia Garrido[62], Pedro Ferreira[58,59,63], Fátima Carneiro[58,59,60].

UMCG: Morris A. Swertz[64], Lennart Johansson[64], Joeri K. van der Velde[64], Gerben van der Vries[64], Pieter B. Neerincx[64], Dieuwke Roelofs-Prins[64].

Charité: Sebastian Köhler[65].

SHU: Alison Metcalfe[42,66].

APHP: Alain Verloes[67,68], Séverine Drunat[67,68].

CHU Bordeaux: Caroline Rooryck[69], Aurelien Trimouille[70]

Telethon UDP: Raffaele Castello[54], Manuela Morleo[54], Michele Pinelli[54], Alessandra Varavallo[54].

Spain UDP: Manuel Posada De la Paz[71], Eva Bermejo Sánchez[71], Estrella López Martín[71], Beatriz Martínez Delgado[71], F. Javier Alonso García de la Rosa[71]

Ospedale Pediatrico Bambino Gesù, Rome: Andrea Ciolfi[72], Bruno Dallapiccola[72], Simone Pizzi[72], Francesca Clementina Radio[72], Marco Tartaglia[72]

University of Siena: Alessandra Renieri[73,74,75], Elisa Benetti[73]

Semelweis University Budapest: Peter Balicza[76], Maria Judit Molnar[76]

University of Ljubljana: Ales Maver[77], Borut Peterlin[77]

University of Lübeck: Alexander Münchau[78], Katja Lohmann[78], Rebecca Herzog[78], Martje Pauly[78]

Val d'Hebron Barcelona: Alfons Macaya[79], Anna Marcé-Grau[79]

Hospital Sant Joan de Déu Barcelona: Andres Nascimiento Osorio[80], Daniel Natera de Benito[80]

University of Freiburg: Hanns Lochmüller[81,82,83], Rachel Thompson[81,83], Kiran Polavarapu[81]

University of Oxford: David Beeson[84], Judith Cossins[84], Pedro M. Rodriguez Cruz[84]

University of Tampere: Peter Hackman[85], Mridul Johari[85], Marco Savarese[85], Bjarne Udd[85,86,87]

University of Cambridge: Rita Horvath[88]

Catalan Institute of Oncology, Barcelona: Gabriel Capella[89], Laura Valle[89]

KU Munich: Elke Holinski-Feder[90], Andreas Laner[90], Verena Steinke-Lange[90]

TU Dresden: Evelin Schröck[91], Andreas Rump[91,92]

[1] Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany.
[2] Centre for Rare Diseases, University of Tübingen, Tübingen, Germany.
[3] Department of Neurodegeneration, Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany.
[4] German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.
[5] Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands.
[6] Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands.
[7] Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.
[8] Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands.
[9] Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands.
[10] Center for Molecular and Biomolecular Informatics, Radboud university medical center, Nijmegen, The Netherlands.
[11] Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands.
[12] Department of Genetics and Genome Biology, University of Leicester, Leicester, UK.
[13] John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research Institute, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.
[14] Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9WL, UK.
[15] Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University Hospitals NHS Foundation Trust, Health Innovation Manchester, Manchester M13 9WL, UK.
[16] Dijon University Hospital, Genetics Department, Dijon, France.
[17] Dijon University Hospital, Centre of Reference for Rare Diseases: Development disorders and malformation syndromes, Dijon, France.

[18] Inserm - University of Burgundy-Franche Comté, UMR1231 GAD, Dijon, France.
[19] Dijon University Hospital, FHU-TRANSLAD, Dijon, France.
[20] Dijon University Hospital, GIMI institute, Dijon, France.
[21] University of Burgundy-Franche Comté, Dijon Economics Laboratory, Dijon, France.
[22] University of Burgundy-Franche Comté, FHU-TRANSLAD, Dijon, France.
[23] CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona 08028, Spain.
[24] Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[25] EURORDIS-Rare Diseases Europe, Sant Antoni Maria Claret 167 - 08025 Barcelona, Spain.
[26] EURORDIS-Rare Diseases Europe, Plateforme Maladies Rares, 75014 Paris, France.
[27] INSERM, US14 - Orphanet, Plateforme Maladies Rares, 75014 Paris, France.
[28] Institut National de la Santé et de la Recherche Medicale (INSERM) U1127, Paris, France.
[29] Centre National de la Recherche Scientifique, Unité Mixte de Recherche (UMR) 7225, Paris, France.
[30] Unité Mixte de Recherche en Santé 1127, Université Pierre et Marie Curie (Paris 06), Sorbonne Universités, Paris, France.
[31] Institut du Cerveau -ICM, Paris, France.
[32] Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres Research University, Paris, France.
[33] Centre de Référence de Neurogénétique, Hôpital de la Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France.
[34] Hôpital de la Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France.
[35] Sorbonne Université, INSERM UMRS_974, Center of Research in Myology, 75013 Paris, France.
[36] AP-HP, Centre de Référence de Pathologie Neuromusculaire Nord, Est, Ile-de-France, Institut de Myologie, G.H. Pitié-Salpêtrière, F-75013 Paris, France.
[37] Institut de Myologie, Equipe Bases de données, G.H. Pitié-Salpêtrière, F-75013 Paris, France.
[38] AP-HP, Unité Fonctionnelle de Cardiogénétique et Myogénétique Moléculaire et Cellulaire, G.H. Pitié-Salpêtrière, F-75013 Paris, France.
[39] Department of Biology and Medical Genetics, Charles University Prague-2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic.
[40] European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.
[41] Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA.
[42] Florence Nightingale Faculty of Nursing and Midwifery, King's College, London, UK.
[43] Genetic Counselling, Genomics England, Queen Mary University of London, Dawson Hall, EC1M 6BQ, London.
[44] MRC Centre for Neuromuscular Diseases and National Hospital for Neurology and Neurosurgery, UCL Queen Square Institute of Neurology, London, UK.
[45] Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK.
[46] Dubowitz Neuromuscular Centre, UCL Great Ormond Street Hospital, London, UK.
[47] NIHR Great Ormond Street Hospital Biomedical Research Centre, London, United Kingdom.
[48] Peripheral Neuropathy Research Group, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium.
[49] Institute Born Bunge, Antwerp, Belgium.
[50] Peripheral Neuropathy Research Group, University of Antwerp, Antwerp, Belgium.
[51] Neuromuscular Reference Centre, Department of Neurology, Antwerp University Hospital, Antwerpen, Belgium.
[52] Laboratory of Neuromuscular Pathology, Institute Born-Bunge, University of Antwerp, Antwerpen, Belgium.
[53] Dipartimento di Medicina di Precisione, Università degli Studi della Campania "Luigi Vanvitelli," Napoli, Italy.
[54] Telethon Institute of Genetics and Medicine, Pozzuoli, Italy.
[55] Unit of Medical Genetics, Department of Medical Sciences, University of Ferrara, Italy.
[56] Institute of Human Genetics, University of Bonn, Bonn, Germany.
[57] Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany.
[58] i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal.
[59] IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Portugal.
[60] Departament of Pathology, Faculty of Medicine, University of Porto, Portugal.
[61] Departament of Genetics, Faculty of Medicine, University of Porto, Portugal.
[62] CHUSJ, Centro Hospitalar e Universitário de São João, Porto, Portugal.
[63] Faculty of Sciences, University of Porto, Portugal.

[64] Department of Genetics, Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.
[65] NeuroCure Cluster of Excellence, Charité Universitätsklinikum, Charitéplatz 1, 10117 Berlin, Germany.
[66] College of Health, Well-being and Life-Sciences, Sheffield Hallam University, Sheffield, UK.
[67] Dept of Genetics, Assistance Publique-Hôpitaux de Paris - Université de Paris, Robert DEBRE University Hospital, 48 bd SERURIER, Paris, France.
[68] INSERM UMR 1141 "NeuroDiderot", Hôpital R DEBRE, Paris, France.
[69] Univ. Bordeaux, MRGM INSERM U1211, CHU de Bordeaux, Service de Génétique Médicale , F-33000 Bordeaux, France.
[70] Laboratoire de Génétique Moléculaire, Service de Génétique Médicale, CHU Bordeaux – Hôpital Pellegrin, Place Amélie Raba Léon, 33076 Bordeaux Cedex, France.
[71] Institute of Rare Diseases Research, Spanish Undiagnosed Rare Diseases Cases Program (SpainUDP) & Undiagnosed Diseases Network International (UDNI), Instituto de Salud Carlos III, Madrid, Spain.
[72] Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, 00146 Rome, Italy.
[73] Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Italy.
[74] Medical Genetics, University of Siena, Italy.
[75] Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy.
[76] Institute of Genomic Medicine and Rare Diseases, Semmelweis University, Budapest, Hungary.
[77] Clinical institute of genomic medicine, University medical centre Ljubljana, Slovenia.
[78] Institute of Neurogenetics, University of Lübeck, Lübeck, Germany.
[79] Neurology Research Group, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain.
[80] Neuromuscular Disorders Unit , Department of Pediatric Neurology. Hospital Sant Joan de Déu, Barcelona, Spain.
[81] Department of Neuropediatrics and Muscle Disorders, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
[82] Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
[83] Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada.
[84] Nuffield Department of Clinical Neurosciences, University of Oxford, UK.
[85] Folkhälsan Research Centre and Medicum, University of Helsinki, Helsinki, Finland.
[86] Tampere Neuromuscular Center, Tampere, Finland.
[87] Vasa Central Hospital, Vaasa, Finland.
[88] Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK.
[89] Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain.
[90] Medical Genetics Center (MGZ), Munich, Germany.
[91] Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, Technical University Dresden, Dresden, Germany.
[92] Center for Personalized Oncology, University Hospital Carl Gustav Carus, Technical University Dresden, Dresden, Germany.