The impact of SOLVE-RD to research & future of RD Diagnostics

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

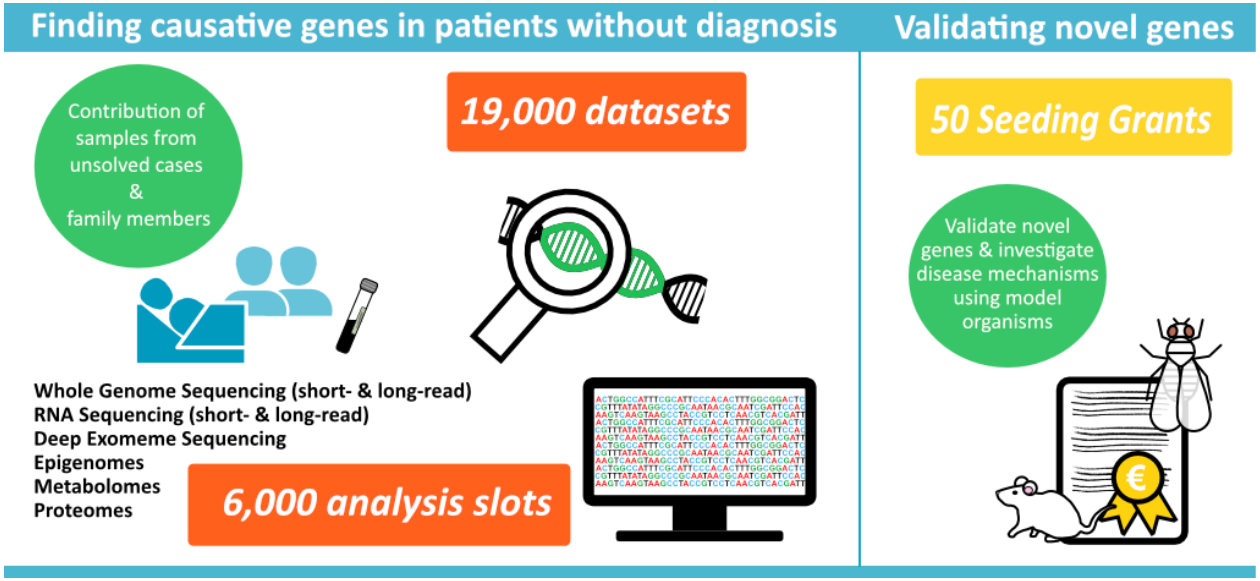
Agenda

- Solve-RD project
- Results (by date)
- Impact (selection)

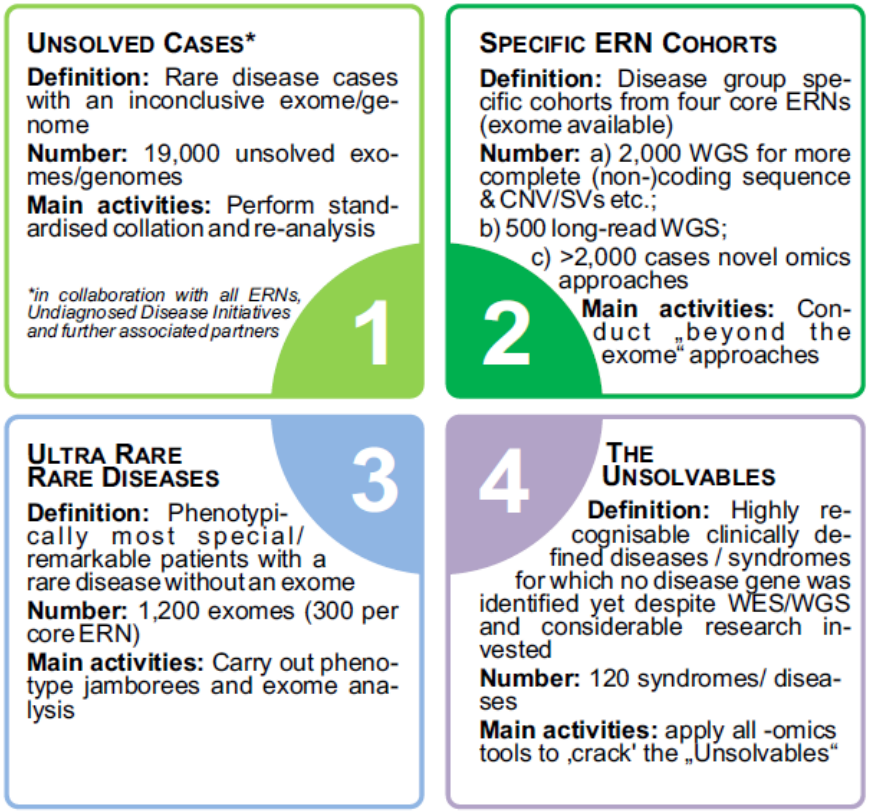
Solve-RD – solving the unsolved rare diseases

- EU funded research project
- 1.1.2018 – 30.06.2023
- 22 partners from 10 countries
- Coordinated by University of Tübingen

One project – one aim



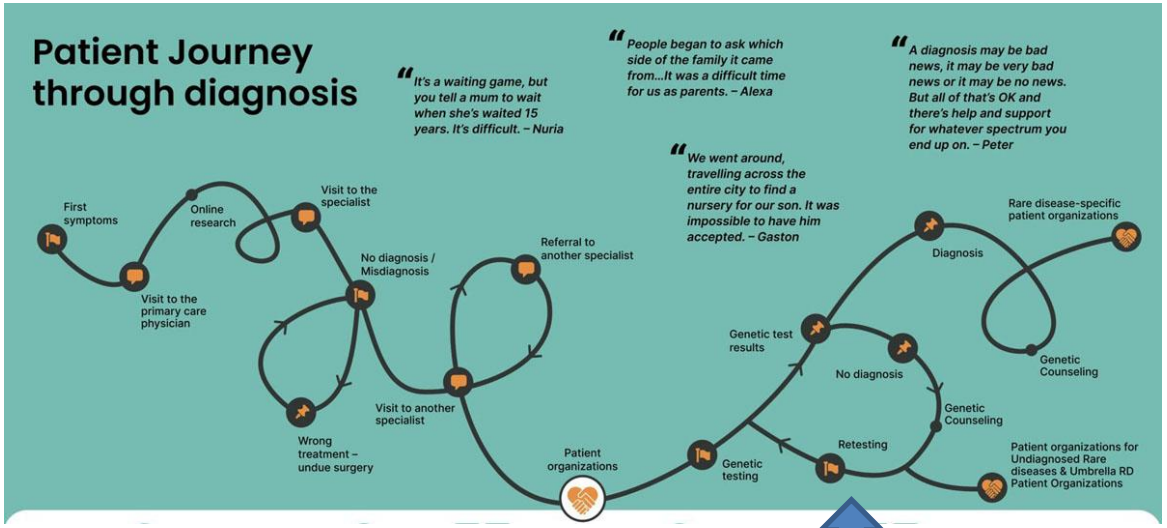
Solve-RD cohorts



Cohort 1: massive data re-analysis

Cohort 2-4: novel combined -omics approaches

Definition of unsolved



- Patients with suspected genetic RD and without a molecularly proven genetic diagnosis
- Patients for which exome sequencing has been done

Results

- Network
- Data re-analysis approach
- Solvathon concept
- Data analysis infrastructure → next talk
- RD data resource

Solve-RD network

- Truly Pan-European project
- ERN based = ITHACA, RND, Euro NMD, GENTURIS
- Newly added ERNs = RITA, EpiCare
- Undiagnosed RD Patient Programme

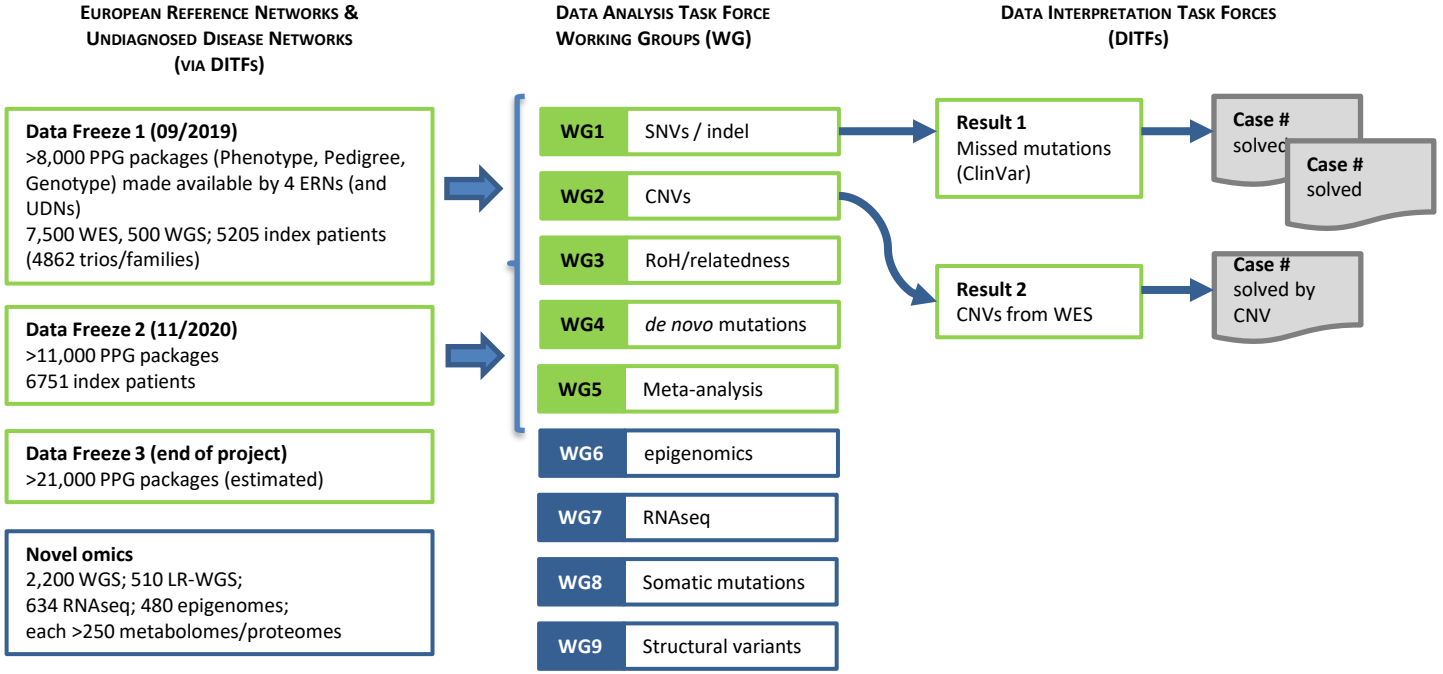
  - 22 beneficiaries
  - 23 associated partners
  - 40 collaborators
  - > 200 groups
  - >300 (clinical) scientists

Re-analysis of previously available data

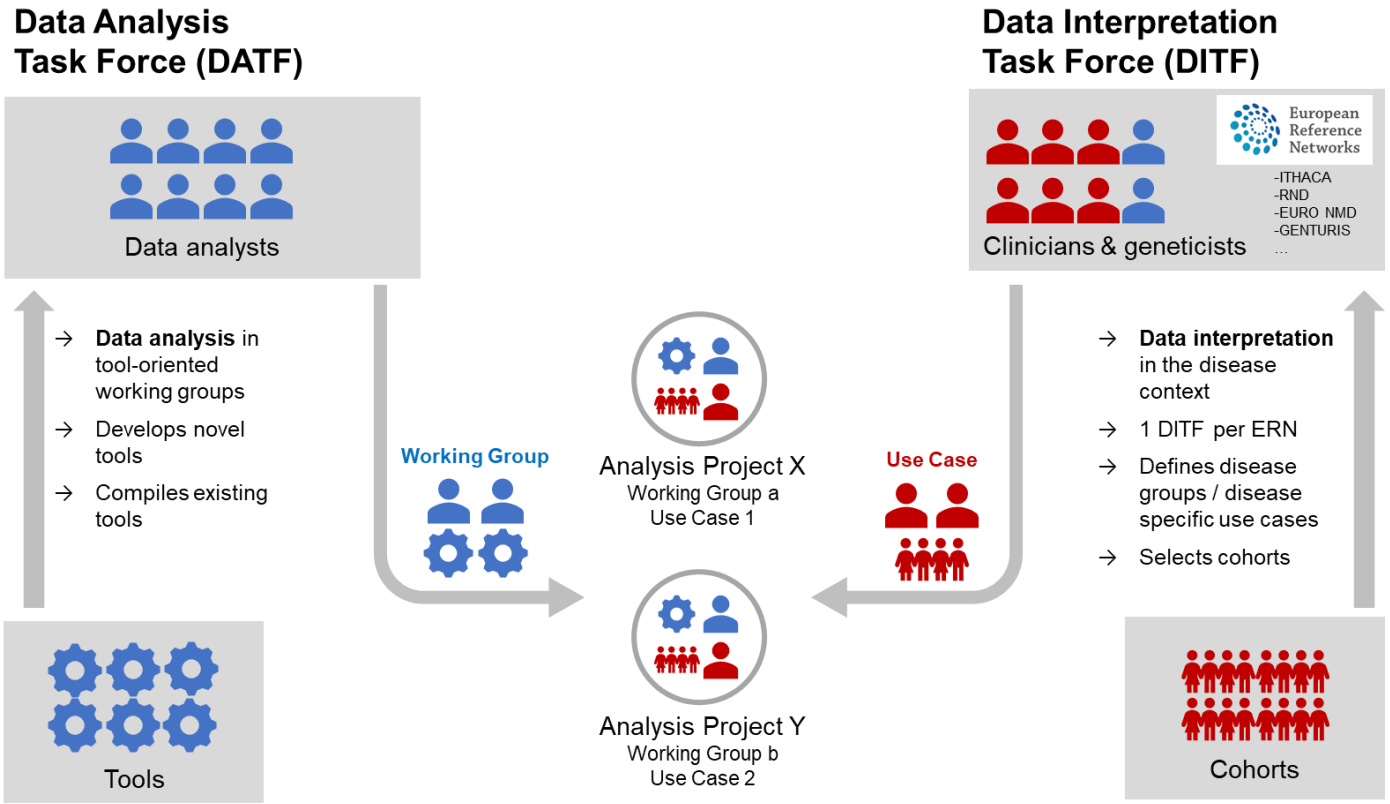
1.) Re-analysis of >23,000 ‘unsolved’ rare disease datasets
- genetic data: exomes, genomes
- pedigree
- phenotypes: HPO, ORDO codes

First demonstration: 6,004 RD cases/families

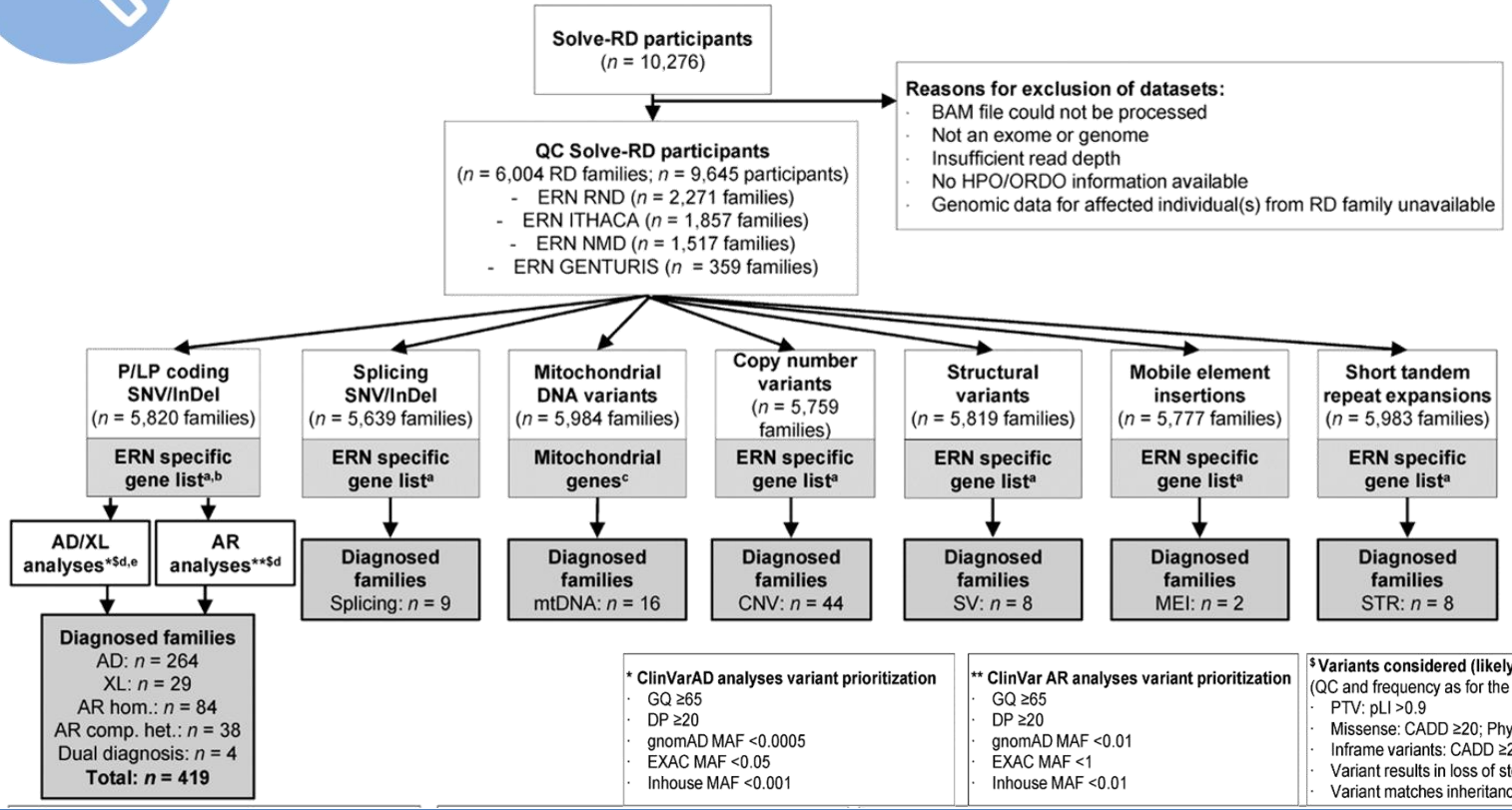
Solve-RD – (re-)analysis approach



Solve-RD data analysis organisation



Systematic re-analysis aligned across 4 diseases



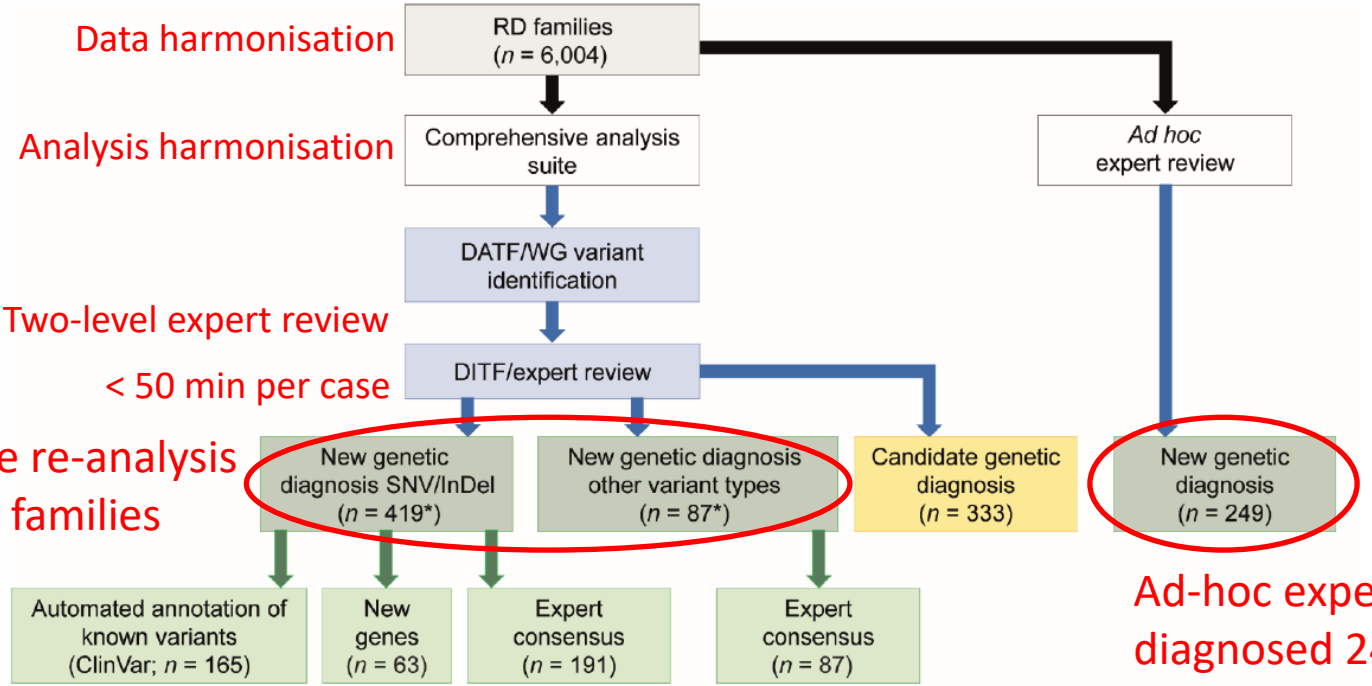
Re-analysis: yield

Data harmonisation

Analysis harmonisation

Two-level expert review

< 50 min per case



Comprehensive re-analysis diagnosed 506 families (8.4%)

Ad-hoc expert review diagnosed 249 families (4.1%)

Re-

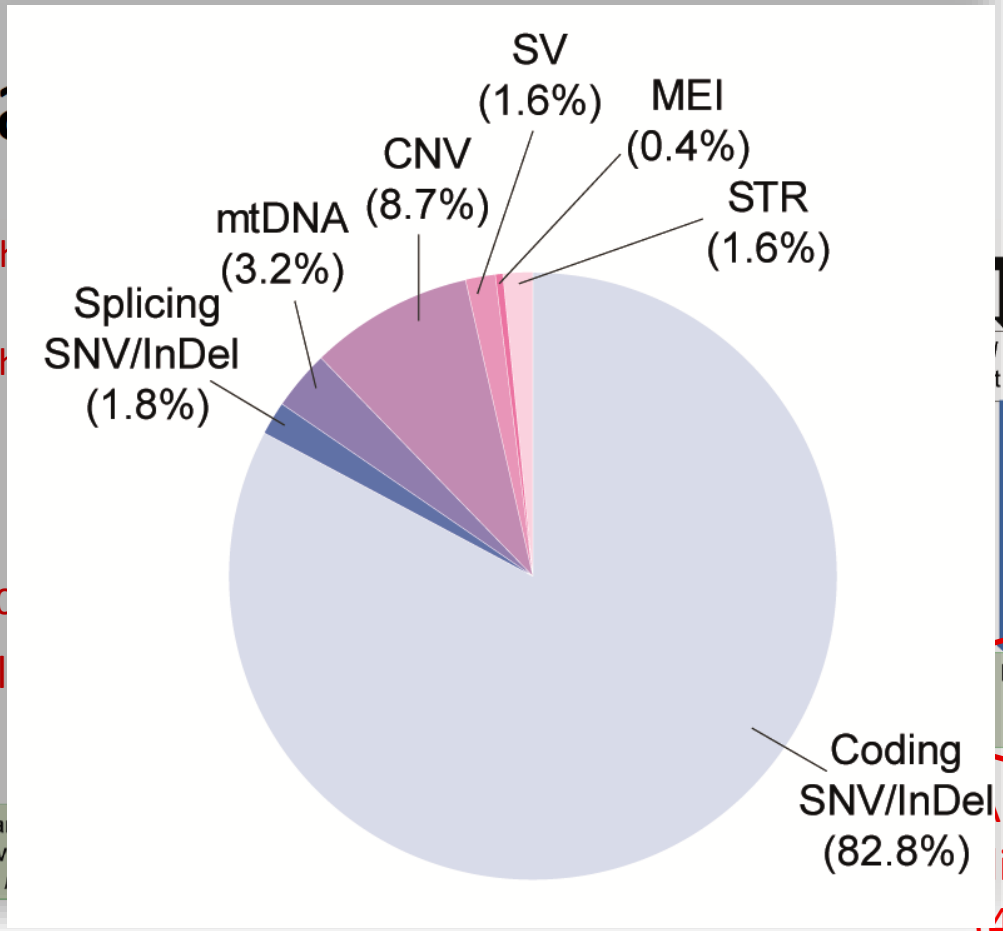
Data

Analysis

Two-level

< 5

Comprehensive re-anal
diagnosed 506 families
(8.4%)



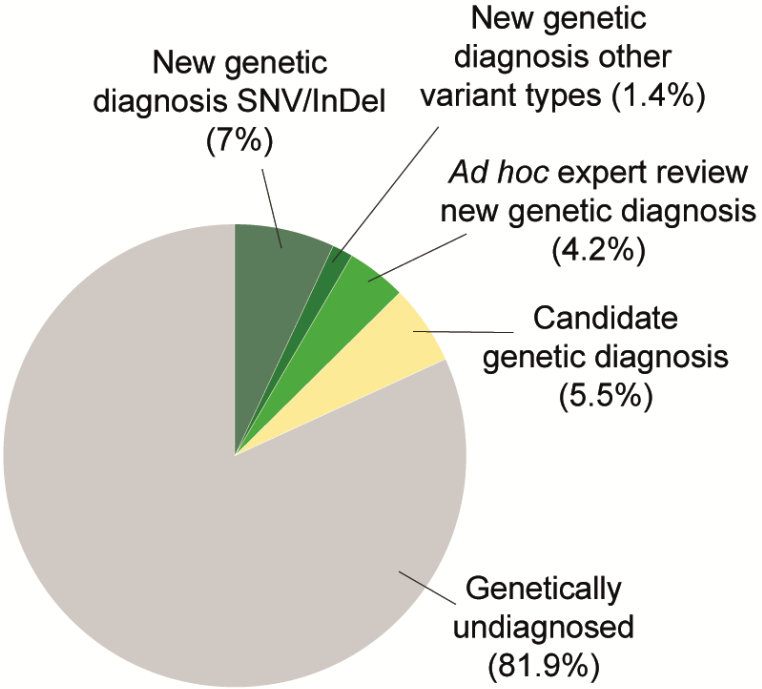
Automated a
known v
(ClinVar;

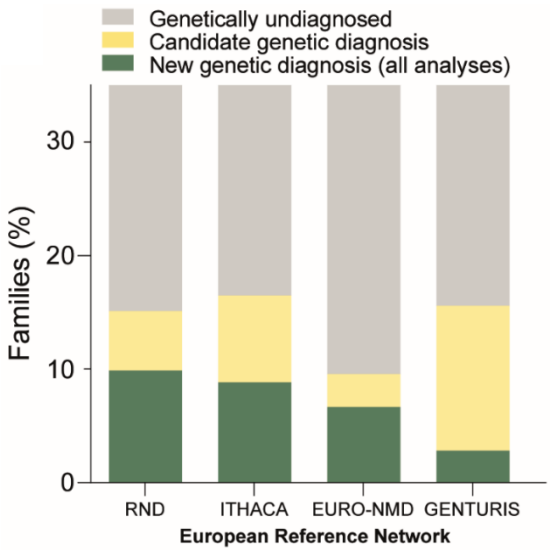
hoc
review

New genetic
diagnosis
(*n* = 249)

d-hoc expert review
iagnosed 249 families
4.1%)

Re-analysis: yield per ERN





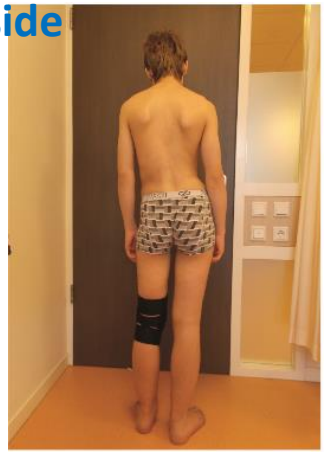
Total <u>families</u> with new diagnosis:

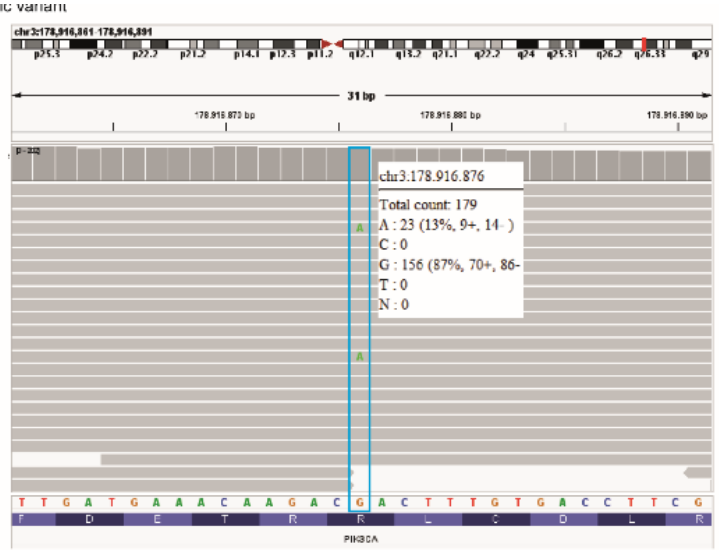
- Systematic re-analysis: 506
- *Ad hoc*: 249
- Candidates: 333

Example of a missed SNV

ERN-ITHACA focuses on rare congenital malformation syndromes and intellectual disability

24-year-old male:
Facial asymmetry, Autism, Seizures, Lower limb asymmetry. Described as **under-development of the left side**





Solve-RD ES Reanalysis:

Detected a variant in PIK3CA in 23/179 reads

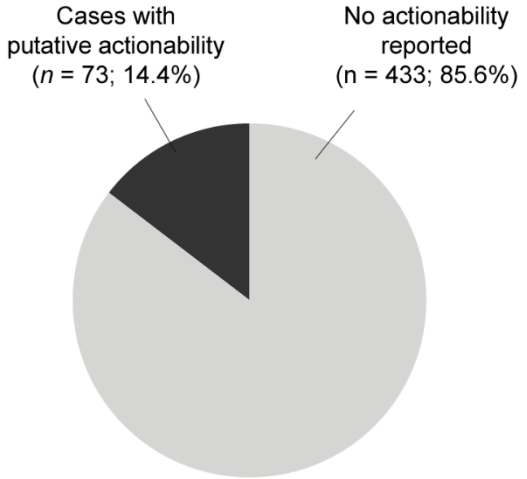
Rare de novo mosaic (13% of the reads) missense variant

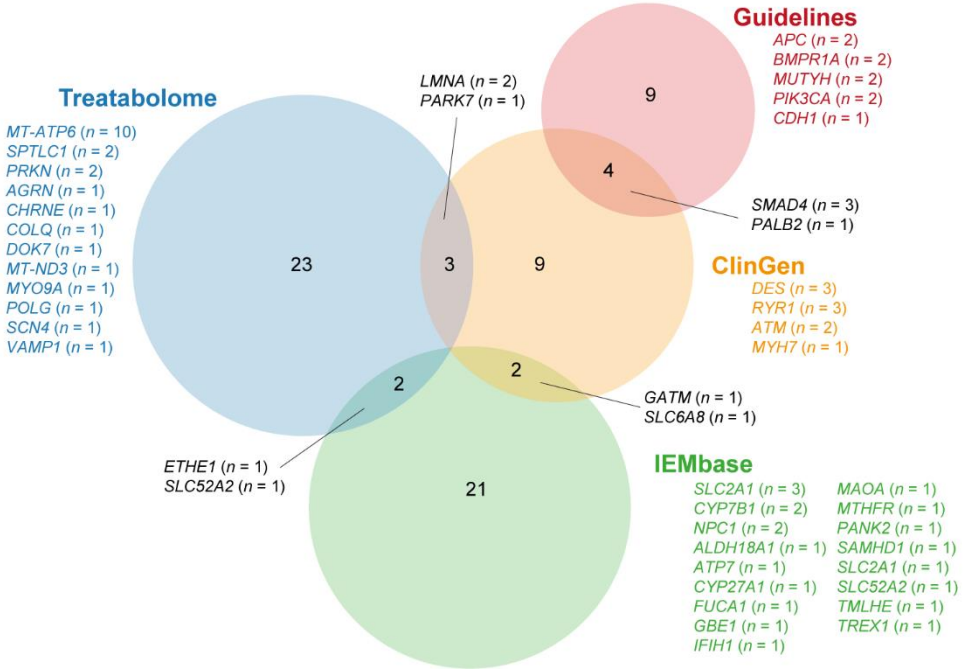
ClinVar Variant – PIK3CA over-growth

Change in clinical diagnosis and resolution of his diagnostic odyssey

Beyond diagnoses

506 newly diagnosed families





Lesson learned

- Even after a decade of diagnostic exome sequencing **new gene-level** or **variant-level information** becomes available
- Analyses are not harmonised across countries/centres; 15.9% (n=87) of novel diagnoses due to **individually rare variant types**
- **Numbers matter: identical variants** in multiple individuals: 21 (likely) pathogenic variants 2(-3) times across 6,004 families (some across ERNs), also increasing global recurrences (*e.g.* ClinVar, Matchmaking)
- **14.4%** of novel diagnoses are amenable to **treatment/actionable → some actions happened**
- Framework & **two-level expert review** are practical blueprint for global scale
- SNV/InDel re-analysis 4.8 minutes per variant, or **42.8 min on average per proband**

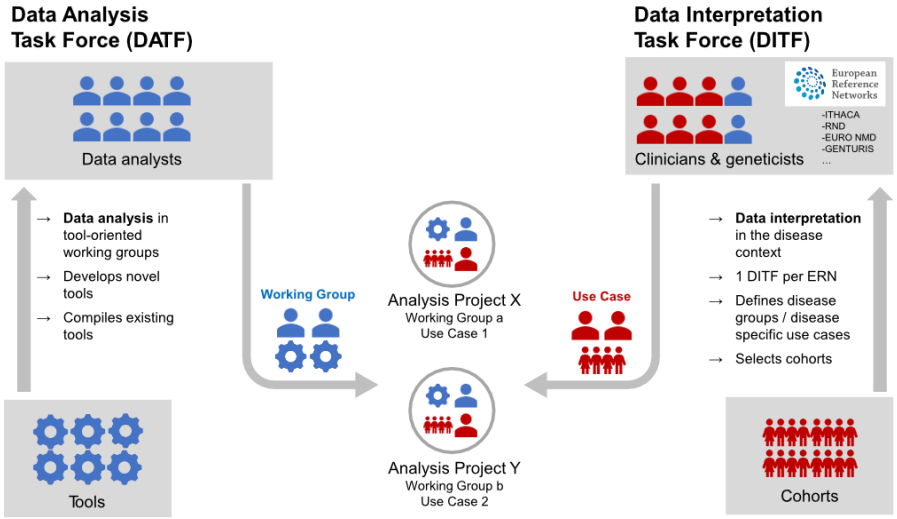
Laurie et al. Genomic Reanalysis of a Pan-European Rare Disease Resource Yields >500 New Diagnoses. Nat Med, accepted.

SOLVE-RD SOLVATHON CONCEPT

Two-level expert analysis and interpretation

1. Bioinformatic and (molecular) genetics experts working together in dedicated working groups within a Data Analysis Task Force (DATF), and

2. Clinical and RD experts from each ERN (DITF) who jointly prioritised and interpreted all variants returned by the DATF



Solve-RD Solvathon concept

1. **Production and release of analysis results (DATF WG)**
2. **On-site or online Solvathon workshop bring together DITF and DATF comprising**
    a. Presentation of analysis concepts and tools as well as analysis results (DATF) and analysed cohort (DITF)
    b. Explanation of interpretation workflow (including required IT setup)
    c. Demonstration of exemplary interpretations of analysis results
    d. Interpretation work in smaller groups (DITF), DATF members assist if needed
3. **Continued interpretation work after the Solvathon workshop**
    a. Based on resources of Solvathon workshop (presentations / recordings / guidance)
    b. Communication between DATF and DITF facilitated through
        i. continuously open communication channel such as Slack or Teams
        ii. defined consultation hours
4. **Interpretation freeze**

Solvathons

workshop bringing together DITF and DATF expertise

Agenda | Solve-RD Solvathon – multi-omics analysis workshop

Wed-Fri, 6-8 March 2024 in Barcelona, Spain, and online | all times CET

Wednesday, 6 March 2024

Venue: *CNAG, Parc Científic de Barcelona (PCB), Calle Baldiri i Reixac, nº4, 08028 Barcelona, Spain [Map] (Room 3, floor -1 (underground) Tower I)*

9:00	**Welcome**	
9:00-9:15	Opening + Welcome (Sergi Beltran, Alex Hoischen)	
9:15-9:45	Epigenetics introduction and cohorts (Julia Schulze Hentrich)	
9:45-10:15	Epigenetics data visualisation / interpretation workflow (Thomas Hentrich)	
10:15-10:45	Epigenetics data analysis approaches (Midhuna Immaculate Joseph Maran and Emanuel Raineri)	
11:00	**Coffee Break**	
11:30-12:00	RNA Seq cohort introduction (Vicente Yépez)	
12:00-12:30	RNA outlier detection (Vicente Yépez)	
12:30-13:00	RNA integration with genotype & phenotype (Vicente Yépez)	
13:00	**Lunch**	
14:00	**demo and hands-on work and interpretation / online consultancy sessions**	
	subgroups per ERN on-site	on-demand online consultancy sessions
15:00	**Coffee Break**	
15:15	**demo and hands-on work and interpretation / online consultancy sessions**	
	subgroups per ERN on-site	on-demand online consultancy sessions
17:00	**End of workshop day 1**	
19:30	**Joint dinner at Citrus Restaurant** *Pg. de Gràcia, 44, 08007 Barcelona, Spain [Map]*	

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

Thursday, 7 March 2024

Venue: *CNAG, Parc Científic de Barcelona (PCB), Calle Baldiri i Reixac, nº4, 08028 Barcelona, Spain [Map] (Room 3, floor -1 (underground) Tower I)*

9:00	**Start of workshop day 2**	
9:00-9:45	Open questions and findings from day 1	
9:45-10:30	Theory with focus on integration (Anna Esteve)	
10:30	**Coffee Break**	
11:00	**demo and hands-on work and interpretation / online consultancy sessions**	
	subgroups per ERN on-site	on-demand online consultancy sessions
13:00	**Lunch**	
14:00	**demo and hands-on work and interpretation / online consultancy sessions**	
	subgroups per ERN on-site	on-demand online consultancy sessions
15:00	**Coffee Break**	
15:30	**demo and hands-on work and interpretation / online consultancy sessions**	
	subgroups per ERN on-site	on-demand online consultancy sessions
16:30	**End of workshop day 2**	
16:30	**Optional tour of the CNAG**	

Friday, 8 March 2024

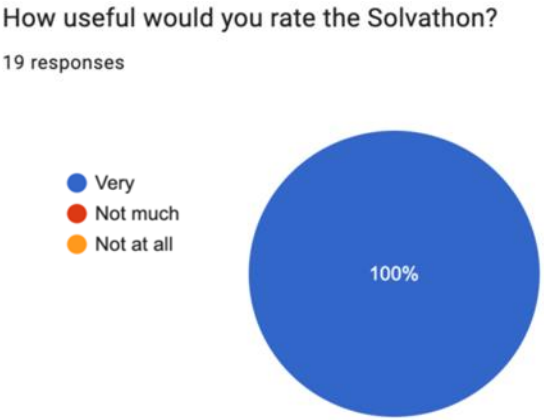
Venue: *CNAG, Parc Científic de Barcelona (PCB), Calle Baldiri i Reixac, nº4, 08028 Barcelona, Spain [Map] (Room 3, floor -1 (underground) Tower I)*

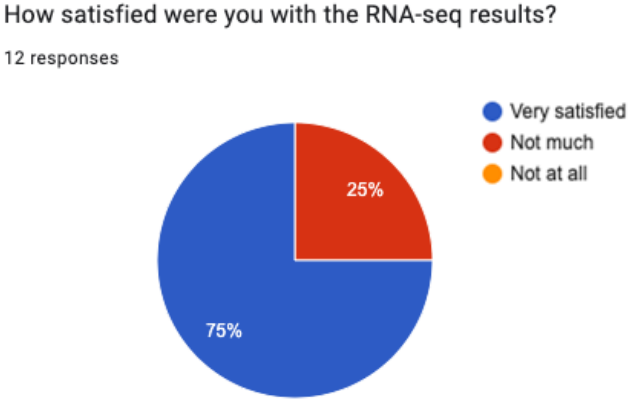
9:00	**Start of workshop day 3**
9:00	**hands-on work and interpretation / preparation of presentations on interesting findings and solved cases**
10:30	**Coffee Break**
11:00	**Presentations & discussion of solved/interesting cases**
12:00	**Wrap up and closing remarks**
12:30	**End of workshop and departure**

Educational aspects

***"Solvathon"* to present and explain results and solve cases**

Solvathon, CNAG, Barcelona, Feb 2023





Solvathon iterations

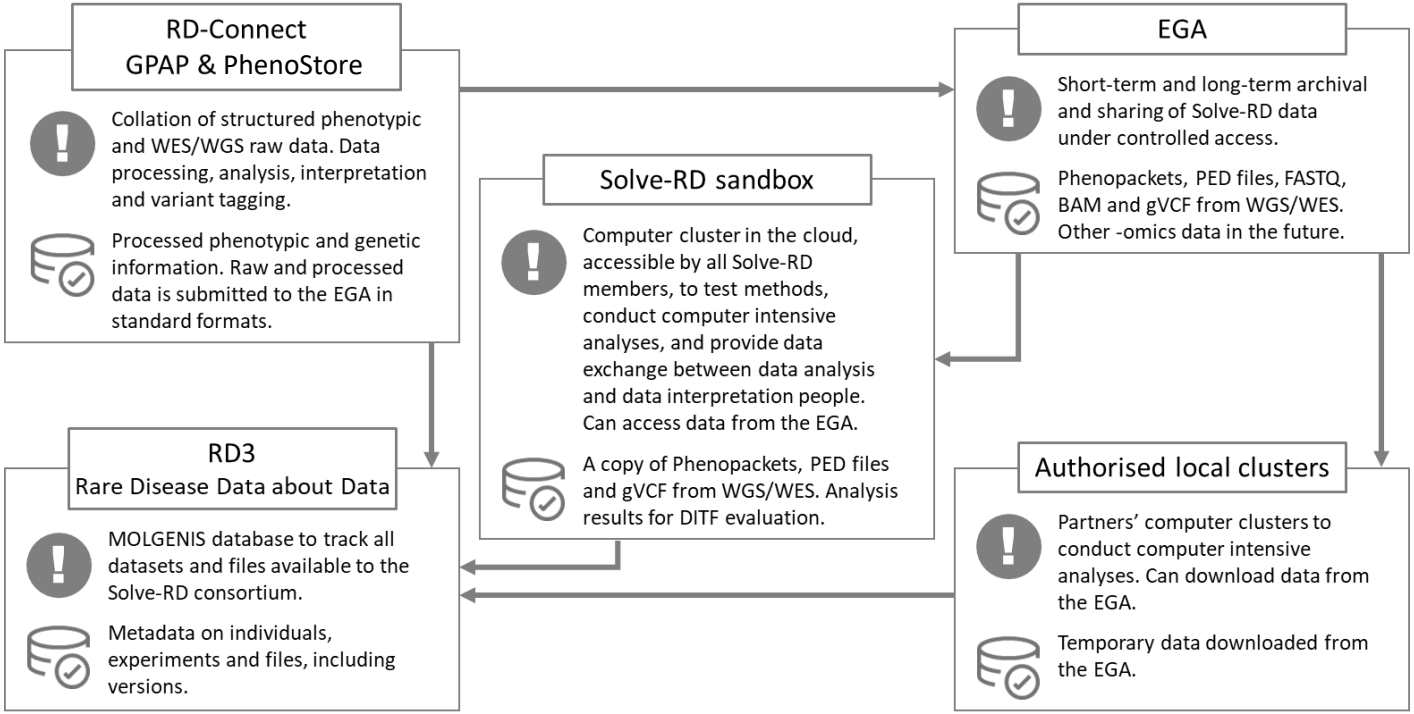
Formats

- ✓ On-site workshop
- ✓ Fully virtual workshop
- ✓ Fully hybrid workshop

Omics / Integration

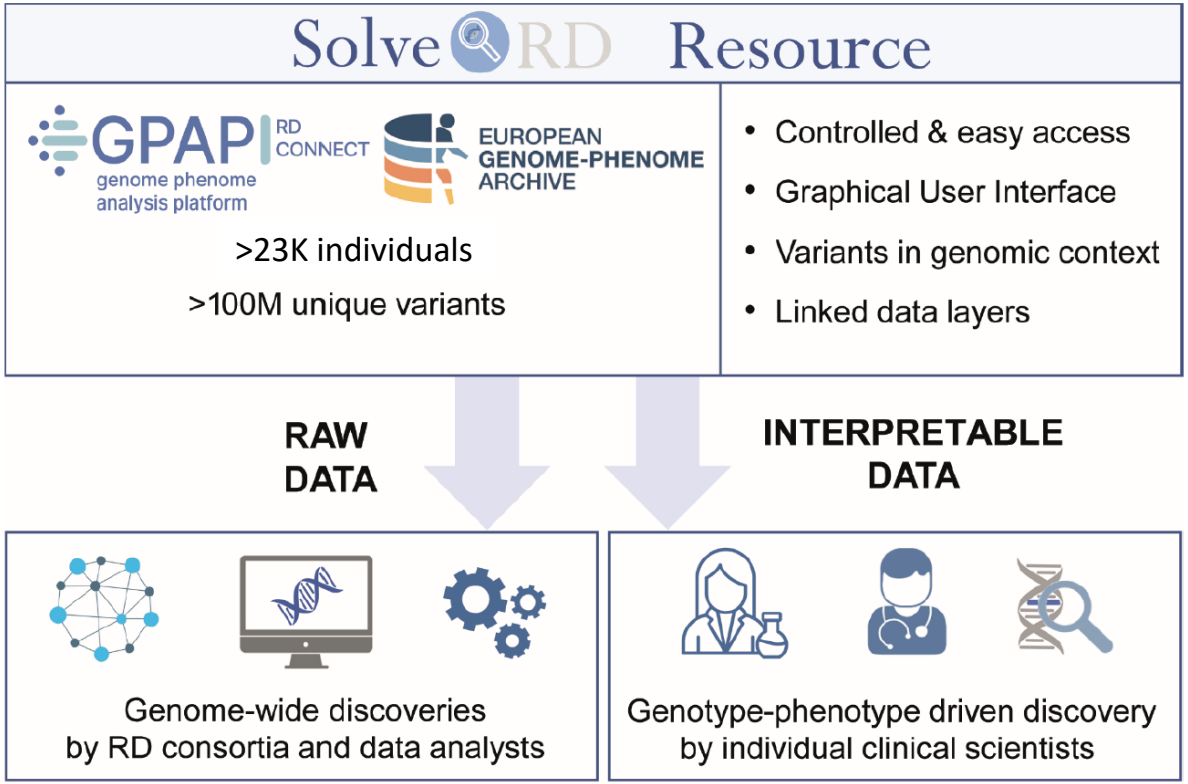
- RNAseq
- SR-WGS
- LR-WGS & OGM
- Multi-omics

Solve-RD data infrastructure



RD-NEXUS (Café Variome): further ERN data discoverability

RD Data Resource for the entire community



Impact

- ERDERA

ERDERA
European Rare Diseases Research Alliance

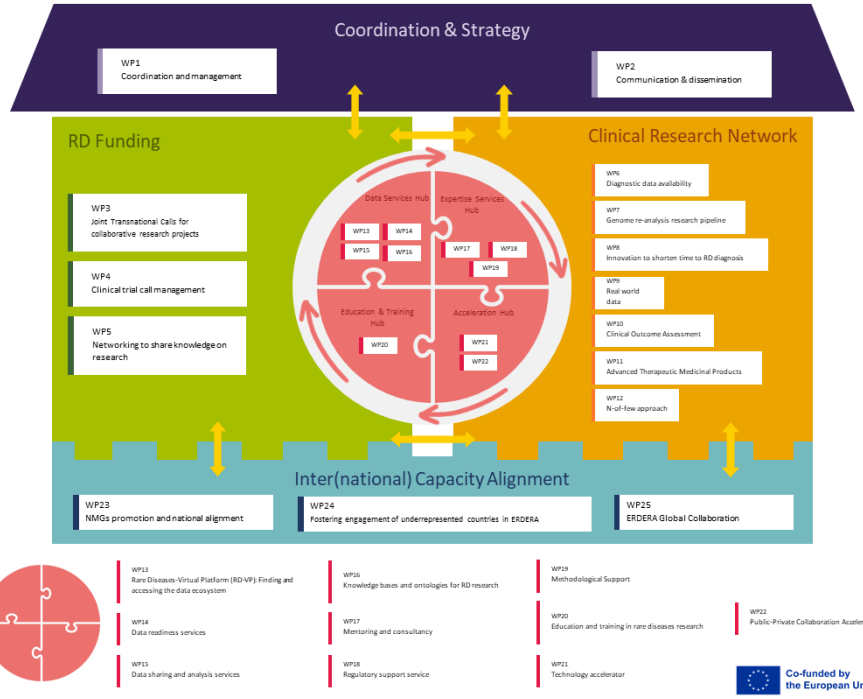
178 Organisations

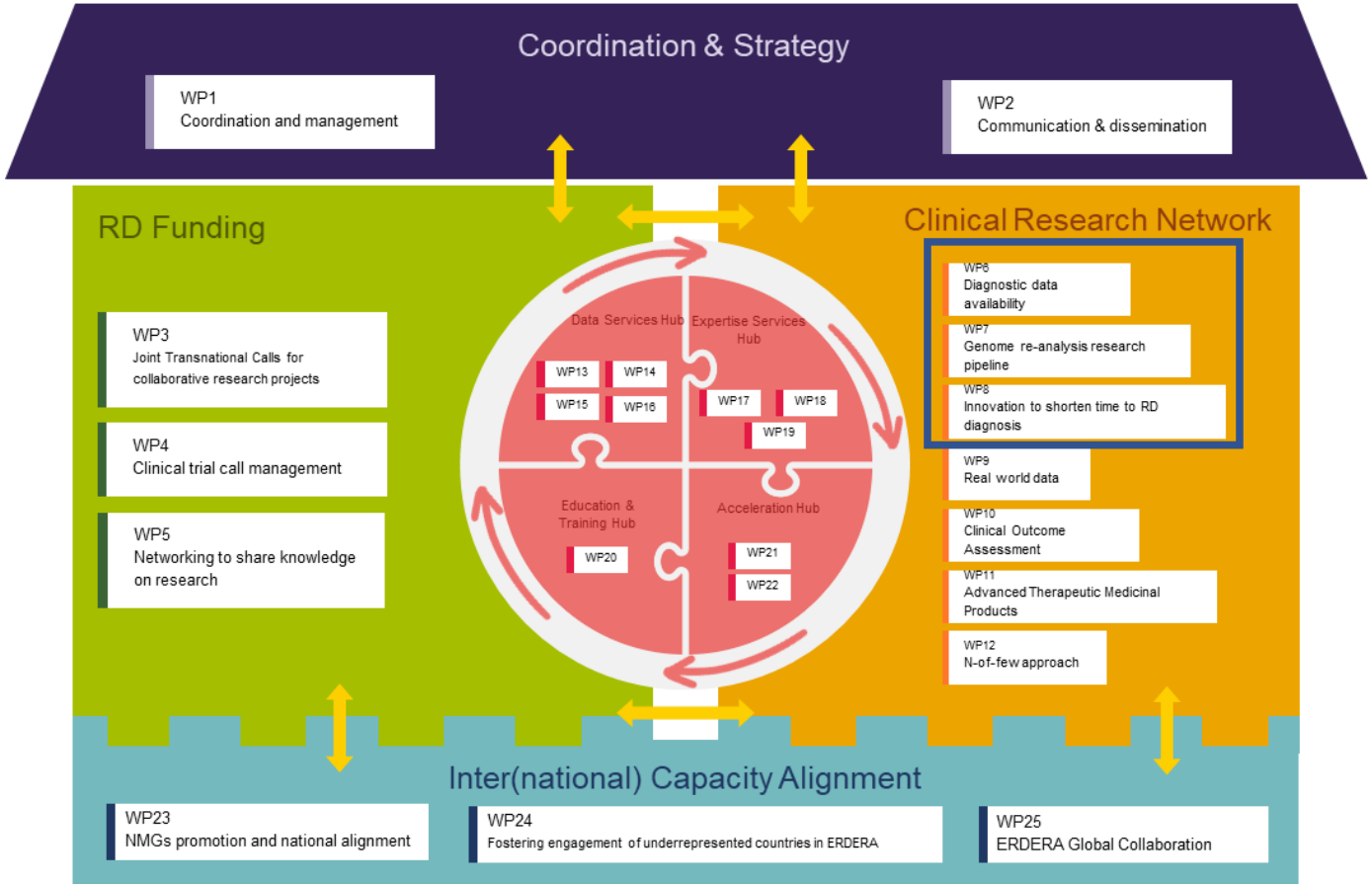
40 funders
81 research performing organisations
9 patients' organisations
3 research infrastructures
22 private for-profit partners (industry & SME)
23 other (univ, hospital, non-profit, public administration)

37 Countries

26 EU member states
8 associated countries
3 non-EU*

*at the time of proposal submission





WP13
Rare Diseases-Virtual Platform (RD-VP): Finding and accessing the data ecosystem

WP14
Data readiness services

WP15
Data sharing and analysis services

WP16
Knowledge bases and ontologies for RD research

WP17
Mentoring and consultancy

WP18
Regulatory support service

WP19
Methodological Support

WP20
Education and training in rare diseases research

WP21
Technology accelerator

WP22
Public-Private Collaboration Accelerator

Main objectives - CRN diagnostic research

WP6	• **Coordinate a pan-European diagnostic research data readiness and collation effort** • Standardization, submission and harmonization of data for reanalysis • Archive the data following the FAIR principles ensuring access to the data during the project and beyond
WP7	• **Coordinate a pan-European genomics diagnostic research process beyond the state-of-the-art diagnostic pipeline** • **Re-analyse existing data through a distributed and federated approach to discover novel diagnoses** • Leverage knowledge and develop best practices to improve variant interpretation across disease groups and ERNs and discover new gene-disease associations • Translate clinical research infrastructure into routine clinical practice
WP8	• Enable **complete genome analysis for RD diagnoses** (for underrepresented countries) to shorten time to diagnosis • New **genomics and integrated multi-omics approaches** to understand/interpret genetic variation enabling improved RD diagnoses.

“Solve-RD 2.0” – bigger, better, closer to healthcare

Further scaling-up from Solve-RD!

Data re-analysis:

1.) From 20k to >100k datasets

2.) From 4-6 ERNs to (almost) all ERNs (n=19)

3.) From big national (rare disease expert) centers, to more/all national centers

4.) From ES to WGS to LR-WGS to –other omics datasets

5.) Scaling up of Solvathon concept

Innovations to shorten time to diagnosis:

1.) Long-read genomics as generic first tier tests

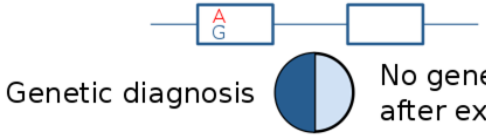
2.) Integration of DNA-RNA tests and pipelines

3.) Multi-omics RD approaches and AI-integration

Many thanks to all!

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

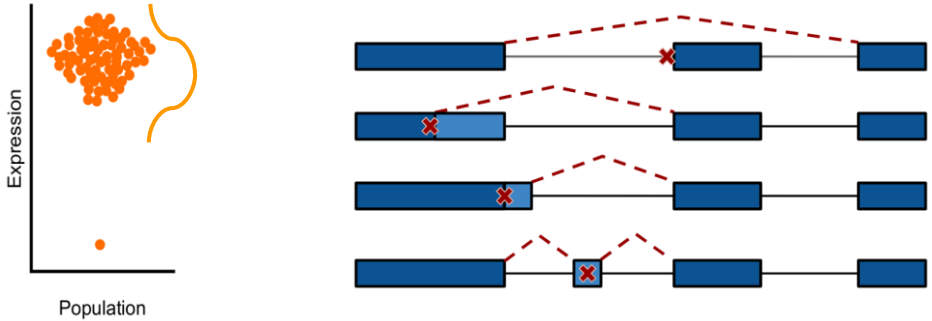
Gene expression to increase diagnostics



1. Sequence RNA (from clinically-accessible tissues)

~15% diagnostic increase over WES

2. Aberrant (not differential!) expression & splicing detection

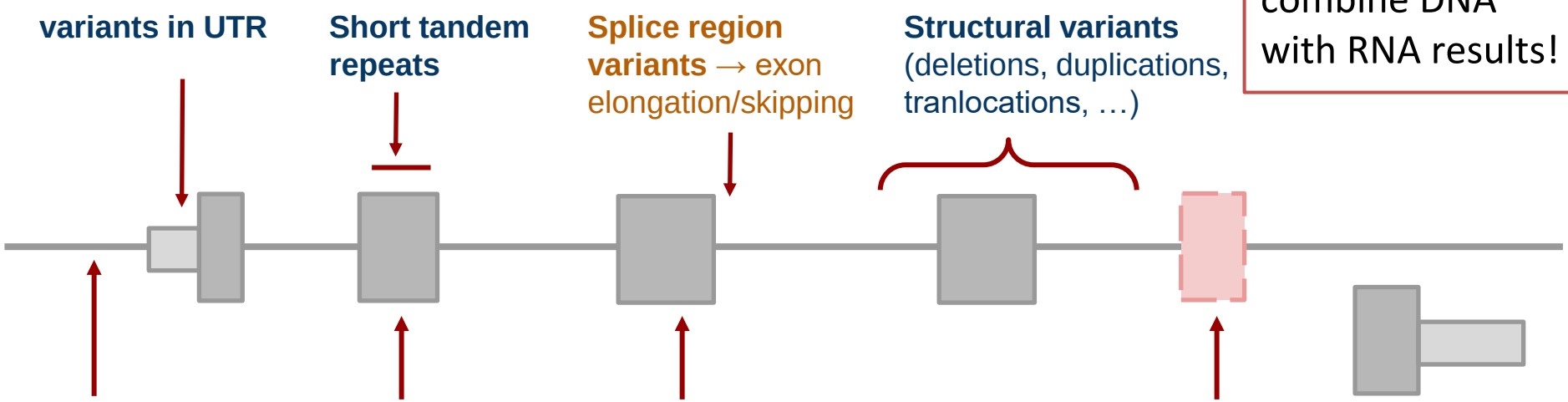


Important to first train the participants!

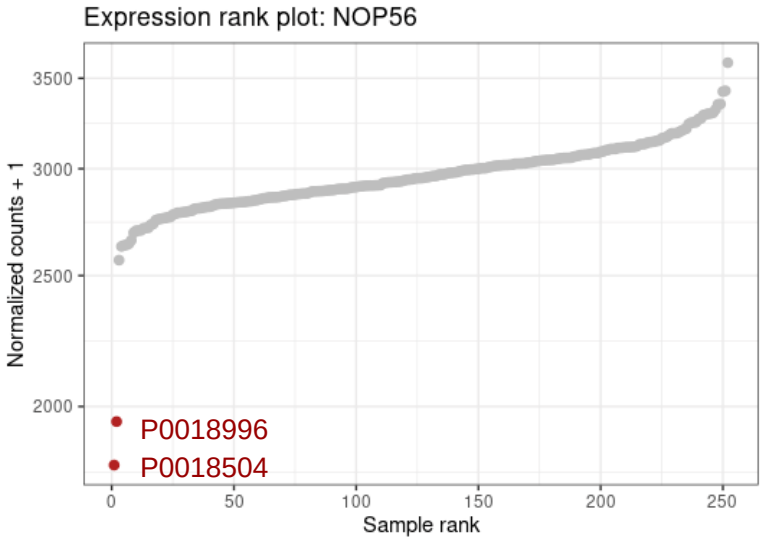
3. Integrate with DNA and clinical data

RNA-seq offers the possibility to inspect the effect of a wide variety of variants

Important to combine DNA with RNA results!



2 RND unrelated patients with aberrant expression in *NOP56*



Importance of multidisciplinary participants!

RNA

- ~50% expression reduction in NOP56 in 2 unrelated individuals

Clinical

- Autosomal dominant inheritance
- NOP56 associated with spinocerebellar ataxia, fitting both patients' phenotypes

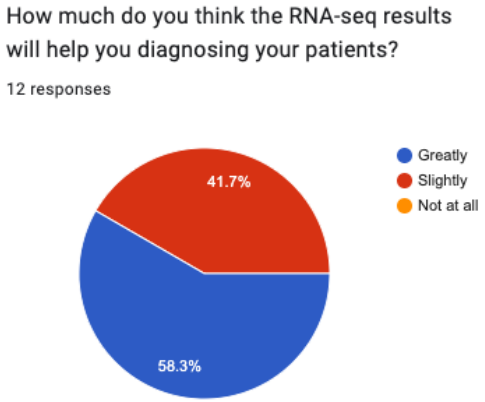
DNA

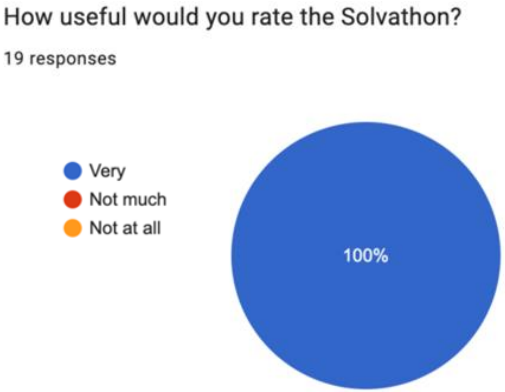
- Standard SNV/indel analysis -> no candidates
- Repeat expansion analysis through ExpansionHunter on site
  - P0018996: >45 repeats
  - P0019023: >34 repeats (father of P0018996)
  - P0018504: >50 repeats

2 solved cases in <2 hours!

Positive feedback!

“All the participants came to solve cases so:
1) it is not passive learning
2) the motivation of people is real, it is not just to learn the abstract things for their future work but to work on their patients
3) there were no need to "fill in the gaps" in schedule, people had enough work to do so everyone was very busy all the workshop”





Lesson’s learned – practical tips

Box 1: Practical recommendations for large scale distributed genomic re-analysis initiatives.

1) **Harmonise pheno-clinical data and metadata**, and make sure it is accessible together with the corresponding genomic data.

2) For heterogeneous collections use **raw sequencing data** as input.

3) Perform **quality control of all data as early as possible** and define strict inclusion criteria. *e.g.* make sure samples are biologically related in the manner described in the phenotypic submission. **Require a minimum on-target coverage** - we recommend **80-fold for ES and 30-fold for GS.**

4) Apply **genome-wide variant calling**, irrespective of enrichment kit used for exome sequencing.

5) Use **multiple variant calling pipelines** for each variant type, with the possible exception of SNV/short InDels, for which variant calling is relatively robust and reproducible. Of all other variant types, **CNVs promise the highest yield from exome data**, as found here and by Lemire *et al* [54]

6) Consider reducing stringency with respect to observed alternative allele frequency for heterozygous calls (*i.e.* allow values below 20%), or apply bespoke **somatic mutation** calling algorithms, if variants are observed in genes commonly associated with the observed phenotype, in order to allow detection of mosaicism or true heterozygotes with poor allele balance.

Lesson’s learned – practical tips

Box 1: Practical recommendations for large scale distributed genomic re-analysis initiatives.

7) Prioritise variants according to their occurrence in **clinical interpretation databases such as ClinVar, HGMD**, and similar local/national resources.

8) **Reverse phenotyping can be key** to re-evaluate the clinical diagnosis in some cases, especially for syndromic disease.

9) Update **bioinformatic workflows regularly to incorporate new tools** and the latest versions of key databases such as ClinVar.

10) If it is necessary to prioritise among cases for re-analysis, **focus first on cases which were (re-analysed) further in the past**, since diagnostic yield is likely to be higher.

11) **Collect feedback** on disease-causing and prioritised candidate variants and solved cases in an accessible database.

12) To facilitate feedback on variant interpretation, **favour specificity over sensitivity**, and share **short-lists** of variants for each individual **once, and only once**.