Deliverable

D3.3 Diagnostic yield per RD per ERN for all omics used in WP2	
Version | Status	V1 | final
Work package	WP3
Lead beneficiary	24 - SRUMC
Due date	31.12.2023 (M72)
Date of preparation	17.04.2024
Target Dissemination Level	Public
Author(s)	Lisenka Vissers (SRUMC)
Reviewed by	Holm Graessner (EKUT)
Approved by	Gisele Bonne (INSERM)

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

Explanation according to GA Annex I:

Evaluate the effectiveness of omics technology at specific RD level (as selected by the ERNs in WP1) by means of diagnostic yield as primary outcome measure.

Abstract:

In the Solve-RD project, seven different -omics technologies have been used to increase diagnostic yield for individuals with a genetically undiagnosed rare disease. These include deep exomes, short-read genomes, long-read genomes and optical genome mapping, transcriptomics, epigenomics, and metabolomics together tailored to not only detect the primary genetic underlying variant, but also to assess functional impact thereof. For each -omics, the ERNs have contributed multiple RD cohorts to unveil the diagnostic potential. In this task, we aimed to determine these yields per RD and per ERN across the omics used.

In total, >3,500 new datasets were generated across the ERNs and -omics, varying between 89 for optical genome mapping and 1,859 short read genomes. So far, diagnostic yield can mostly be obtained per technology across ERNs as interpretation is still ongoing. Highest diagnostic yield has been obtained for short read and long read genome sequencing and optical genome mapping, at 8%, 11% and 12%, respectively. Yet, when also taken into account ERN specific subcohorts, transcriptomics for individuals with clinical characteristics reminiscent of a metabolic disease (ERN-ITHACA, RND and Euro-NMD) and amplicon-based long reads sequencing for Lynch(-like) syndrome and transcriptomics (ERN-Genturis) provide most new diagnosis (at 16% and 19%, respectively). Interestingly, for few cases, it has been demonstrated that a combination of multiple omics was need to find a diagnosis, underscoring the complexity of RD diagnosis.

In conclusion, each novel omics technology for which data have been interpreted contributed to the new diagnoses of individuals with genetically unexplained rare disease. With analysis and interpretation efforts still ongoing, it remains to be established which diagnostic test serves which patient cohort, or ERN best.

Introduction

The aim of this deliverable was to determine the diagnostic yield per RD per ERN across the various omics technologies used in WP2. Establishing a genetic diagnosis in a research environment is crucial for progress, but is only one important step in the overall pathway for a patient with a rare disease. For the work done in Solve-RD to have maximum impact on patients' lives, it must become integrated into healthcare systems in a way that meets patient needs and that forms the most streamlined and cost-effective way of providing a diagnosis. In addition, receiving a diagnosis must not be seen as the end goal but rather as the starting point enabling the most appropriate treatment pathway for the patient. This deliverable focuses on translating outcomes to the clinic and to patient benefit by assessing the diagnostic yield obtained via the novel omics in WP2 across the four ERNs and for specific RD subtypes which received focus in the Solve-RD project.

Report

Sample collection and data production has been a continuous effort throughout the project. Analysis and Interpretation of the data was organized via DATF/DITF structures in WP2. For

most omics, data analysis is still ongoing, and feedback on diagnostic yield is so far limited to specific -omics and subtypes of RD within ERNs. An overall summary based on the current information is presented in Table 1.

Table 1 | Overview of -Omics samples and individual solved status (RD cases analysed within Solve-RD).

Omics	**# Samples (analyzed/ total collected)**	**ERN** *(n = families included)*	**Cases solved**
Deep exome	425/620 (215 tumor samples analysed)	ITHACA (n=29) RITA (n=9) NMD (n=3) Genturis (n=384)	tbd tbd tbd 11 cases
Short read genome	1,859/1,943 (842 completed families)	ITHACA (n=114)* RND (n=274) NMD (n=243) Genturis (n=102) EpiCare (n=93) RITA (n=16)	4 (4%) 27 (10%) 36 (15%) - (0%) 4 (5%) - (0%)
Long read genome	445/650 (180 completed families)	ITHACA* (n=42) RND (n=52) NMD (n=60) EpiCare (n=26)	1 case 6 cases** 3 cases -
Optical Genome	168/216 (84 completed families)	ITHACA* (n=57) RND (n=8) NMD (n=13) RITA (n=6)	2 cases 5 cases*** 1 case 2 cases
Metabolome	36/91	ITHACA* (n=19)	Interpretation to start
Transcriptome	528/573	ITHACA* (n=41) RND (n=140) NMD (n=103) Genturis (n=13) EpiCare (n=42)	- 9 cases** 7 cases - 2 cases
Epigenome	146/411	ITHACA* (n=27) RND (n=13) NMD (n=8) Genturis (n=2)	Interpretation to start

**ITHACA submitted only cases/families with unsolvable syndromes for these omics. It is expected that an X-omics interpretation is required to find the genetic diagnosis. To determine diagnostic yield, ITHACA cases were therefore excluded from the denominator. ** For two cases, the X-omics approach of GS and RNAseq led to the discovery of the diagnosis. ***strong genetic evidence but pending functional validation of disease mechanism.*

Diagnostic yield from overall -omics technologies per ERN and RD as a whole, are yet to be determined. Yet, in addition to the search for generalizable statements, Solve-RD has also provided detailed per -omic for RD subtypes. A selection of these are provided below:

Short read genome sequencing for NDD (ERN ITHACA)

In has been well reported that genome sequencing (GS) is able to provide a genetic diagnosis for patients who remained undiagnosed after having reached the 'end of routine care', often including exome sequencing. One of the ERN-ITHACA partners tested whether GS is a better first-tier genetic diagnostic test than current exome-based standard of care (SOC) by assessing the technical and clinical validity of GS for patients with neurodevelopmental disorders (NDD) (PMID 36114283). They performed both GS and exome sequencing in 150 consecutive NDD patient-parent trios, with the primary outcome being diagnostic yield, calculated from disease-causing variants affecting exonic sequence of known NDD genes. GS (30%, n=45) and SOC (28.7%, n=43) had similar diagnostic yield. All 43 conclusive diagnoses obtained with SOC testing were also identified by GS. SOC, however, required integration of multiple test results to obtain these diagnoses. GS yielded two more conclusive diagnoses, and four more possible diagnoses than ES-based SOC. Interestingly, these six variants detected only by GS were copy number variants (CNVs). These data demonstrated the technical and clinical validity of GS to serve as routine first-tier genetic test for patients with NDD. Although the additional diagnostic yield from GS is limited, GS comprehensively identified all variants in a single experiment, suggesting that GS constitutes a more efficient genetic diagnostic workflow.

Added diagnostic yield GS for NDD when compared to ES, when interpretation is limited to coding sequence:

1-2% conclusive diagnosis

3-5% possible diagnosis

Long Read Genomes for Rare Disease:

Solve-RD aims to identify the disease-causing genetic variant in previously undiagnosed rare disease (RD) families; So far, Solve-RD conducted one of the largest long-read sequencing (LRS) studies to date. We evaluated the effectiveness of LRS in resolving genetic causes in RD families lacking a clear molecular diagnosis despite previous testing. We studied 293 individuals from 114 genetically undiagnosed RD families selected by ERN experts. The 93 families in the 'unsolved' cohort were affected by rare neurological, neuromuscular, or epilepsy disorders, and 21 families by clinically recognizable 'unsolvable syndromes' for which genetic causes remain unknown. Using 10x coverage HiFi-LRS we detected small variants, SNVs, structural variants (SVs), and short terminal repeats (STRs), with *de novo* analysis in trios. Results: Variant analysis in known disease genes resulted in novel genetic diagnoses in ten families with neurological or neuromuscular disorders, involving *de novo* and inherited SVs, SNVs, and STRs in know disease genes. Additionally, we found potential disease-causing SVs in three families. In an 'unsolvable syndromes' patient, we identified a known disease-causing SNV in *TUBA1A*, re-classifying the patient's diagnosis. While we did not identify a common genetic cause in any of the 'unsolvable syndromes', we identified causal genetic variants in 11% of families from the 'unsolved' cohort, and candidate variants in an additional 3%. Our study shows the potential and effectiveness of even modest-coverage LRS in solving RDs.

Overall added diagnostic yield Long Read Genomes for unsolved Rare Disease:

11% conclusive diagnosis

3% possible diagnosis

Optical Genome Mapping for Rare Disease

Optical genome mapping (OGM) allows genome-wide detection of structural variants (SV) as small as 500 bp. Here, we systematically applied OGM to identify disease-causing SVs in rare disease (RD) patients that remained undiagnosed after standard-of-care analysis in the respective RD-expert centers. For 159 individuals from 83 families, with 98 affected individuals' genomes were successfully analyzed using OGM. These included 40 singletons, 11 duos, 26 parent-child trios with expected *de novo* origin of disease, and 6 other family constellations. The largest groups of samples were selected by clinical experts of ERN-ITHACA (57 families, of which 25 trios), followed by 13 NMD-, 8 RND-, and 6 RITA- families. Ongoing data analysis revealed 6 *de novo* SVs in 26 trios (SV mutation rate: ~1⁄4 genomes), confirming an expected low SV mutation rate, but also implying every *de novo* SV as a potential disease candidate. Additionally, we focused on rare (segregating) SVs in singleton cases (30-50/genome). In total, we identified (candidate) disease-causing SVs in >10% of index cases. This includes SVs in well-established disease genes (e.g. *SERPING1* in two hereditary angioedema families), as well as rare inversions (n=1), complex chromosomal aberrations (n=1) or overlapping non-coding SVs (n=5) each suggesting novel disease mechanisms. These SVs were usually missed by the initial test modalities or were identified in parallel by Solve-RDs short- and long-read genome sequencing efforts. This large-scale pan-European study successfully demonstrates the identification of so far hidden SVs in RD by OGM.

Overall added diagnostic yield Optical Genome Mapping for unsolved Rare Disease:

12% (likely) conclusive diagnosis

8% possible diagnosis

Transcriptome analysis for Rare Disease

In 40-60% of individuals with a rare disease, the genetic cause remains elusive. In a substantial group of these unsolved cases, a variant of unknown clinical significance can be identified. Functional follow-up is often required to understand the impact of such variant at transcript and/or protein level. For assessing the impact on transcripts, the use of transcriptomics might be useful. In Solve-RD we generated transcriptomes for 573 samples. Whereas interpretation is still ongoing, a diagnosis has already been established in 9 cases for ERN RND, 7 for NMD and 2 for EpiCare. Interestingly, for at least 2 cases in RND, a cross-omics approach was required to identify the genetic cause, underscoring the value of integrated analysis as neither single omic would have led to the diagnosis.

Overall added diagnostic yield Transcriptomics for unsolved Rare Disease:

At least 5% conclusive diagnosis (uncorrected for number of families)

In addition, to the collective efforts, novel insights for dedicated RD cohort per technology were obtained through collaborative efforts of Solve-RD partners within the ERN.

Transcriptomics for metabolic disorders (ERN Metab, with RD overlap ERNs ITHACA, Euro-NMD, and RND)

The DATF-WG lead on transcriptomics has in parallel to Solve-RD participated in the evaluation of the clinical use of transcriptomics for metabolic disease (PMID 35379322). Whereas ERN-Metab was not specifically involved in Solve-RD, their data are taken into account because of phenotypic overlap of rare disease spectra of ITHACA, NMD, and RND, as well as their experiences given involvement of the DATF-WG lead. Transcriptomics was performed on fibroblasts in a series of 205-WES negative indexes. They detected on average 12,500 genes per sample including around 60% of all disease genes—a coverage substantially higher than with whole blood, supporting the use of skin biopsies. Genes were prioritized based on aberrant expression, aberrant splicing, or mono-allelic expression. The pipeline required less than 1 week from sample preparation to result reporting and provided a median of eight disease-associated genes per patient for inspection. A genetic diagnosis was established for 16% of the 205 WES-inconclusive cases. Detection of aberrant expression was a major contributor to diagnosis including instances of 50% reduction, which, together with mono-allelic expression, allowed for the diagnosis of dominant disorders caused by haploinsufficiency. Moreover, calling aberrant splicing and variants from RNA-seq data enabled detecting and validating splice-disrupting variants, of which the majority fell outside ES-covered region, thus requiring GS for the identification of the genetic variant at genomic level.

Added diagnostic yield Transcriptomics for suspected metabolic disease compared to ES:

16% conclusive diagnosis

(Amplicon-based) long read sequencing for Lynch syndrome (ERN-Genturis)

Long read sequencing (LRS) provides immense advantage over short read sequencing for variation detection, amongst which, the better detection of (noncoding) structural variation. As such, this technology has the potential to replace all other diagnostic workflows for germline testing including NGS based assay, but also cytogenetic testing. In Solve-RD we have merely focused on the added diagnostic yield compared to other diagnostic tests. Whereas LRS can be performed at genome-wide level, for some RD entities with limited genetic heterogeneity, also a targeted LRS approach is useful. In Solve-RD, this strategy was employed for Lynch-syndrome by ERN-Genturis.

Lynch syndrome (LS) is characterized by the development of mismatch repair–deficient (dMMR) colorectal and endometrial cancers at a young age in life. LS is caused by germline

pathogenic variants in 1 of the mismatch repair genes MLH1, MSH2, MSH6, or PMS2 or deletions affecting the EPCAM. Current germline diagnostics for LS include targeted short read NGS and MLPA of the coding regions of the MMR genes and somatic MLH1-promoter hypermethylation. In the absence of a germline PV in an MMR gene, the presence of somatic dMMR is investigated. However, a proportion of individuals with dMMR tumors remain genetically unresolved after germline and somatic analyses. These individuals have an unexplained dMMR tumor and are known as individuals with Lynch-like syndrome (LLS). One of the Solve-RD partners investigated whether a targeted LRS approach would help to uncover diagnostically relevant variants in LLS families. In a cohort of 32 individuals with an unexplained dMMR cancer, they identified at least 6 individuals (19%) with a noncoding diagnostically relevant variants, including both SNVs as well as structural variants. In addition, for 2 families, variants of unknow clinical significance where identified, possibly enhancing added diagnostic value by another 6%.

Added diagnostic yield Transcriptomics for LLS families compared to ES:

19% conclusive diagnosis

6% possible diagnosis

Conclusion

Despite that the analyses and interpretation efforts are still ongoing, the novel omics approaches have already provided valuable insights into the diagnostic yield. Especially the generic approach of long read sequencing (at genome wide level) may provide a successful first tier test, especially when combined by the use of a generic functional follow-up such as transcriptomics.