Co-funded by the European Union
Grant Agreement n°874719

Additional information on the PIN

- extended information about the technical challenges to be addressed per lot

Contents:

The main technical challenges to be addressed for **lot 1 (pre-sequencing)** are:

Pre-sequencing is the initial part of the workflow, which is key for success in integrating the entire NGS workflow into routine diagnostic use. It includes **pre-analytics** (collection of the specimen, storage, transport and extraction of nucleic acids (NA)), and **library preparation and target enrichments**. The main challenge to be addressed is the preservation of the native NA profiles as they were in the patient's body, in a standardized, controlled and documented way, to ensure reliable downstream NGS results, representative of the actual condition present in the patient. Standardization of pre-analytics will need to adhere to the EU IVDR, new ISO/IS and CEN/TS standards which describe requirements for pre-analytical handling and preservation of analyte quality.

For optimal linking of the pre-analytical workflow steps with the library preparation, QC of the extracted nucleic acid must be implemented including the assessment of the integrity and unbiased NA profile, required eluate conditions (e.g. pH, salt concentrations), required NA quantity and concentration, among others.

Library preparation and target enrichment include multiple individual steps (NA fragmentation, amplification, clean up and normalization), and are highly challenging for the reliability and reproducibility of sequencing data by NGS. Therefore, verification and risk mitigation according to the IVDR requirements must be introduced in these steps, focusing on the impact of NGS test results. One important quality control measure is for example the introduction of so-called "molecular barcode" technologies in a gene-specific, primer-based target enrichment process, which assures high analytical sensitivity and specificity for NGS of actionable variants. Another important quality control step during library preparation will be assessment of fragment size distribution and concentration.

The target enrichment strategy will depend on the workflow to be developed (eg. targeted panel-based or whole genome / whole exome – based or other). Implementing a capture-based strategy or other adequate technical solutions is required for the design

of new target panels. As the bioinformatics approach used when dealing with capture panels differs from that used in amplicon-based enrichment strategies, the use of a single enrichment approach would also allow implementing a unique bioinformatics pipeline, hence contributing to the simplification and uniformity of the analytical process. The ability to reliably detect variants with low variant allele frequencies (Vaf) is mandatory when dealing with cancer samples, therefore, Unique Molecular Identifier (UMI, also known as random molecular barcodes) technology should be implemented. In the case of target panels, the providers of clinical panels should be able to update them following a defined update procedure, to keep up with the dynamically evolving list of genes and variants relevant for cancer diagnostics and treatment decisions.

Specimens produced according to relevant standards (e.g., FFPE, fresh tissue, FNA, whole blood, plasma, etc.) will be provided to the operators and for testing the workflows by the buyer's group.

The main technical challenges to be addressed for **lot 2 (sequencing)** are:

The actual sequencing part of an integrated NGS workflow very much depends on the properties and specific requirements of the technology and instrument provider. One of the main technical pitfalls is the accurate connection of the biological process of nucleotide binding/depletion to generation of a chemical or physical signal for transfer into sequence information. Another part which is essential for verification / validation is reproducibility and precision – an important requirement of the EU IVDR. The used NGS technology platform should maintain the required sensitivity and precision, and avoid sequencing bias for reporting analyte profiles as they were in vivo.

The specific issues to be taken into account while developing the sequencing part of the workflow are for example:

- The cost of the sequencing infrastructure, initial installation and staff training, and initial sequencing practice runs.
- Some NGS instruments are more prone to sequencing errors than others, e.g. in microsatellite regions that are very important for clinical cancer sequencing.
- The use of technologies that produce short reads vs long reads can lead to variability in coverage of variants, especially for highly repetitive regions and structural variants, and thus in the interpretation of results and implementation of clinical interventions.
- Some NGS instrument manufacturers are not able to offer sufficiently fast consulting/repair/replacement services.
- Most instruments and kits are "for research use only" and not supported long-term. This sometimes results in kits that have to be expensively validated by the user for clinical testing.

The main technical challenges to be addressed for **lot 3 (bioinformatics analysis)** are:

All components of **data management**, **transfer** and **analysis** throughout the entire workflow process (bioinformatics, in-silico data modelling, e-reporting) must be verified to reduce complexity for clinicians as well as users, and to fulfil regulatory requirements (EU IVDR, FDA etc.), but at the same time avoid analysis bias and false information. The bioinformatics part of the workflow should ensure that sequencing data is processed, filtered, analysed and displayed in the best possible way so that the final NGS information reflects the real molecular status of the patient. The specifications of clinically relevant cut-offs for the intended NGS tests and their verification will be another key element as well as the specification and development of robust calling algorithms.

For each disease, the business processes and metadata schemas should be defined following the entire clinical pathway from the medical specialist's request for genomic analysis inside a patient's electronic health record (EHR), the laboratory workflow (standard operating procedure - SOP), bioinformatics pipeline and returning results back to the patient's EHR where they will be visualized in textual and graphical form. The necessary SOPs and bioinformatics pipelines should be updated when needed, according to clinical requirements and state-of-the-art bioinformatics knowledge.

Processing of patient data has to meet data protection regulation (GDPR). Dedicated infrastructures will have to be available for storage of large amounts of NGS data and backup of data should be automated to prevent data loss. The lack of standardized bioinformatic protocols, coupled with a significant complexity in the current pipelines requires the development of fully validated, user-friendly tools dedicated to the analysis of targeted panels or WES data, supporting complete versioning and annotation lots.

The main technical challenges to be addressed for **lot 4 (integrated reporting)** are:

Translating NGS results into medical decision-making reports, by integrating NGS results with pharmacogenomics panels and existing e-medication tools containing dosing and drug interactions, should be achieved in this lot. Since this information has to be made available to healthcare professionals and patients at the bedside for rapid interpretation, it will be important to determine the optimal method to clearly present NGS results and their medical relevance. A report format which provides the relevant information, enables appropriate interpretation (based on the scientific and clinical evidence) of NGS results and provides decision support for the recommended intervention is going to be crucial. The relevant clinical information should be reported in a concise and clear way, reporting only data with validated evidence for clinical decisions and in a form minimizing the risk of data misinterpretation. Moreover, links to rapid learning tools will also be provided, to facilitate the gradual upskilling of health care practitioners in pharmacogenomics. As patients are often not able to understand the genetic results presented to them, a specific communication plan should be developed focusing especially on the communication of uncertainties that could stem from the NGS results. Strategies to support families in coping with genetic information and to manage barriers related to the disclosure of genetic information within families need to be developed.