Lighthouse project

## Development and validation of a multi-omics workflow to diagnose rare disease

<div class="6u 12u$(small)">  
 <h3>Life-threatening Phenotypes</h3>

Rare diseases are a group of over 7,000 disorders that have a strong impact on health and have major social and economic consequences. They affect 400 million people worldwide, including about 500,000 in Switzerland. Rare diseases often affect children, causing premature death or chronic disability. They are found at increased rates in critically ill neonates or children admitted to ICUs. Rare diseases can cause life-threatening conditions such as metabolic dysfunction, encephalopathy, severe infection, or cardiac failure. Extreme phenotypes (severe forms of rare diseases) are interesting targets for finding rare, causal genetic and biochemical variants because patients are more likely to carry harmful low-frequency or de novo alleles. The identification of causal genetic and biochemical variants in rare, severe diseases can be extremely valuable for patients, particularly children. By gaining a better understanding of these diseases, healthcare professionals may be able to develop more effective treatments and provide better care for affected individuals.

Promise of Multi-omics for Personalized Medicine

In the past decade, the discovery rate of the causes of rare diseases has been increasing due to the ability to generate and analyze genomic data at reasonable cost. A pioneering project in Australia used rapid DNA sequencing (WES) and resulted in a genetic diagnosis in 51% of children, leading to a change in management in 76% of patients. Adding transcriptomic data (RNA-seq) increases the diagnostic yield of WES or WGS. Multi-omics technologies simultaneously assess the genome, transcriptome, proteome, and metabolome, which provides a better understanding of genotype-phenotype relationships. Integration of multi-omics data has the potential to identify rare genomic variants and their functional validation, leading to a higher diagnostic yield compared to WGS alone. Multi-omics has the potential to be a breakthrough diagnostic approach in children with extreme phenotypes by enabling the prompt initiation of lifesaving treatments and/or a change in management.

## Gaps in current practice

WES and WGS are used in diagnostics, but patient selection is biased and there are still a high proportion of patients with undiagnosed diseases. The identification of large numbers of Variants of Unknown Significance (VUS) in DNA sequencing datasets makes the picture more complicated. There is a need for a multi-omics workflow to provide better diagnostic yield compared to WGS.

Current gaps in knowledge and practice include:

1. Delayed diagnosis of rare diseases. There is often a delay in diagnosing rare diseases and inconsistencies in investigating and treating them.
2. Lack of novel approaches to rare disease diagnosis. Conventional methodologies (WES and biochemical testing) are used, but they can be time consuming and have low diagnostic yield.
3. Unavailability of clinical multi-omics. Multi-omics technologies are currently limited to research settings.
4. Clinical diagnostic workflows not adapted for multi-omics. Clinical diagnostic workflows are not designed for discovering novel diseases.
5. Lack of information from end-users on multi-omics. Information on how providers and families respond to multi-omics results is currently unknown.

## Project phases

☑ **Phase 1 Complete:** Multi-omics approach was used to diagnose and characterize 210 individuals affected with the rare inborn error of metabolism methylmalonic aciduria. The combination of genomic, transcriptomic, proteotyping, and metabolomic information with phenotypic dataset led to the identification of a molecular diagnosis for 84% of the patients. The multi-omics technology was found to be powerful for diagnosis and discovery of pathomechanistic factors in rare diseases. A workflow for data acquisition, transfer, storage, and analysis within the PHRT/SPHN infrastructure was established. Another genomic and transcriptomic workflow was established to identify 2 novel diseases in 89 patients with extreme immunological phenotypes. This resulted in identifying rare diseases in children with life-threatening extreme phenotypes through combining clinical evaluation and multi-omics exploration.

☐ **Phase 2 Ongoing:** The integrated approach of WGS, RNA-seq, proteotyping, and metabolomics will be applied to existing cohorts of extreme phenotypes to identify rare deleterious genetic variants. Novel computational methods will be developed to integrate multi-omics data into a diagnostic workflow.

☐ **Phase 3 Ongoing:** The workflow developed in Phase 2 will be tested in a prospective cohort of 140 children with extreme phenotypes. Provider and family uptake of multi-omics information will be examined. The aim is to identify rare diseases causing extreme phenotypes even in single patients.

## Research Design

### Patient Recruitment and Phenotype Selection (Responsible: Group 1)

**Phase 2:** Access existing cohorts of 250 children with extreme phenotypes where consent to use stored biospecimens is available. Children with severe metabolic dysfunction (n=150): (i) Biochemically diagnosed patients without molecular data (ii) Patients with VUS (Variants of Uncertain Significance) (iii) Undiagnosed patients. Children suspected of primary immunodeficiency (n=100): (i) Severe infection in the absence of clinical risk factors (ii) Severe immune dysregulation (Pediatric Inflammatory Multisystem Syndrome associated with SARS-CoV-2).

**Phase 3:** Multi-site study in a prospective cohort of 140 patients admitted to or referred to intensive care. Blood samples (PAXgene for RNA extraction, EDTA for DNA extraction, leukocytes/serum for proteome/metabolome assessment), skin biopsies, and mucosal swabs from family members will be obtained where feasible. A panel of consortium members will review and prioritize cases. Inclusion criteria: (i) Acute encephalopathy (ruled out trauma, infection, intoxication, and asphyxia). (ii) Life-threatening metabolic derangement. (iii) Unusually severe infection. (iv) Primary cardiac failure. (v) Out-of-hospital sudden infant death or near-fatal event. (vi) Unusually severe inflammation syndromes (cytokine storm syndrome, hemophagocytic lymphohistiocytosis, COVID-19-associated Pediatric Inflammatory Multisystem Syndrome). (vii) Undiagnosed diseases (if not already included in Phase 2).

### Clinical Data Processing (Responsible: Group 2)

Implement core SwissPedData set to capture clinical features and outcomes. Test SwissPedData with EHR data, SRDR and MDSi to characterize phenomes. Embed the lighthouse project within SwissPedHealth to: (i) Gather essential clinical data of the included patients. (ii) Enrich SwissPedData with MDSi and SRDS. (iii) Demonstrate how NDS (National Data System) can be enhanced with high resolution omics data. (iv) Establish a whole-of-life data stream to capture future outcomes. (v) Ascertain the incidence of identified rare diseases.

### Multi-omics Data Generation (Responsible: Group 3)

* Phase 2 and Phase 3: Patient-derived Biospecimens: Patient-derived biospecimens from Phase 2 and Phase 3 will be sent to the Swiss Multi-Omics Center (SMOC).
* Nucleic Acid-Based Analyses:

1. WGS and RNA-seq data will be generated and analyzed by the Health 2030 Genome Center (GC).
2. The unaligned DNA and RNA sequence data is clinical grade and can be used for clinical decision making.

* Proteomic Profiling:

1. Performed by the Clinical Proteotype Analysis Center (CPAC).
2. Uses next-generation proteomics technology (DIA-MS).
3. Converts a clinical specimen into a digital proteotype map.

* High-Throughput Targeted/Untargeted Metabolomics and Lipidomics Analyses:

1. Performed by the Clinical Metabolomics Analysis Center (CMAC).
2. Analysis of primary metabolites and xenobiotics is done by high-throughput flow injection, time-of-flight mass spectrometry.

* Data Processing and Sharing:

1. All centers are part of the Swiss Multi-Omics Center (SMOC).
2. SMOC is integrated into the SPHN/BioMedIT secure computational infrastructure.
3. Provides secure data processing and sharing.

### Primary Data Analyses (Responsible: Group 4)

Automated DNA variant calling and annotation to search for novel disease-causing variants. Prioritization of rare variants based on: (i) Potentially damaging effect. (ii) Known function of the gene in the clinical presentation. (iii) Degree of purifying selection. RNA-seq data analysis for quantitative expression profiling and alternative splicing analysis. Coordinated integration and visualization of DNA variants, RNA transcripts, proteoforms, and metabolites. Mapping onto biological networks and pathways using tools developed by SMOC and the Fellay group at EPFL. Data storage and analysis performed within SPHN BioMedIT infrastructure to ensure privacy, security, and adherence to FAIR principles. Downstream analyses using multimodal machine learning to integrate multi-omics and clinical data.

### Multi-omics results using multimodal machine learning analysis (Responsible: Group 5)

**Phase 2:** A machine learning approach will be used to analyze multi-omics data and clinical phenotype data. A cohort of 250 children with extreme phenotypes and fully annotated multi-omics data will be used to train a multimodal model. Multi-view clustering will be used to identify features that are most important for the prediction. New methods suited for rare disease diagnostics will be developed for data with a low sample size but high dimensionality. A novel probabilistic approach to multimodal constrained clustering will be derived to guide the algorithm towards a desirable configuration. The validated machine learning workflow will be used to identify relevant features within a complex multi-omics dataset.

**Phase 3:** The validated machine learning workflow will be utilized in a prospective cohort. The workflow will be optimized for singleton diagnosis using simultaneously obtained whole blood specimens. The model output will provide a probability scale of multi-omics data explaining the observed extreme phenotype, aligned with ACMG recommendations. Enrichment with trio or kindred WGS data, or cell-derived functional datasets will be possible.

### Implementation of findings (Responsible: Group 6)

Study information will be fed back from the research setting into clinics. Diagnostic yield will be defined as the primary outcome. The study results will be reviewed by an expert panel. Information on potentially relevant genetic findings will not be released if not related to a rare disease causing the extreme phenotype. Collaboration with the undiagnosed disease network and care4rare will occur where appropriate.

### Bioethical acceptability and implementation from end-users (Responsible: Group 7)

**Phase 1:** Literature Review and Clinical Experiences. Review of the literature on attitudes towards the offer and return of results from WES or WGS in the rare disease setting in countries outside of Switzerland. Clinical experiences of associated investigators Ormond and Rauch (medical geneticist University of Zurich) who have been providing predictive and diagnostic genetic testing results for decades. PPI-focused discussions with patients and family members who have undergone previous clinical omics testing.

**Phase 2:** Exploratory Interviews. Up to 25 medical providers and 25 parents will be interviewed in English, German or French to determine their attitudes towards a hypothetical offer and return of results. The interview guide will be developed based on the findings from phase 1. Interviews will be transcribed, coded, and analyzed.

**Phase 3:** Identification of Information Needs and Best Practices. Identify the information families need to consider in order to make informed decisions about the returned results. Determine how best to facilitate discussions with families and medical providers. Identify a process to return results that maximizes descence and autonomy and minimizes risks of potential harm. Nest this work with the PPI work done as part of SwissPedHealth WP 4102.

## Scientific Output and Expected Contributions

Development of a Novel Multi-Omics Workflow

The lighthouse project will deliver a multi-omics workflow optimized for rare disease diagnosis. This workflow goes beyond the current state-of-the-art of WES/WGS and will deliver genetic, molecular, and functional information.

Identification of Novel Diagnoses and Pathomechanisms

Multi-omics analysis of existing and prospective cohorts will identify novel diagnoses and pathomechanisms for rare diseases. Machine learning will be implemented to integrate the data and demonstrate its usefulness as a diagnostic tool.

Standardized Framework for Rare Disease Diagnostics

The provider and patient perspective will be included to provide a standardized framework for all pediatric Swiss research hubs. The prospective cohort will focus on extreme phenotypes predominantly recruited in intensive care settings, but the approach is scalable to other conditions and settings. The multi-omics workflow will be combined with the SwissPedHealth dataset to obtain information on incidence, clinical criteria, and outcomes. The SwissPedHealth data stream will be used to test data-driven strategies to pre-select patients most likely to benefit from multi-omics. The lighthouse project will demonstrate the capacity of the SwissPedHealth data stream to integrate research projects with high density omics data.

Accelerate Discovery in Rare Disease Diagnostics and Therapeutics

The goal of the lighthouse project is to set a sustainable framework for translational studies on children with the aim to accelerate discovery in rare disease diagnostics and therapeutics.