Multi-omics

This page describes how we generate data for our patients. We provide some examples of how the data is analysed and list short examples of the types of software required and how we process data.

## Data generation

We generate data using the Swiss Multi-Omics Center (SMOC) [http://smoc.ethz.ch](http://smoc.ethz.ch/). This data is then analysed according to the personal requirements for each patient. We use the Swiss Multi-Omics Center (SMOC) molecular profiling platform to provide us with robust data, research, and simultaneous insights for clinicians. This allows use to generate better informed treatment decisions based on molecular insights.

“Each patient’s individual molecular makeup and unique characteristics will be the basis for guiding the medical decisions and restoring health.” Personalized Health and Related Technologies (PHRT) is a strategic focus area of the ETH Domain; the latter embracing key research institutions like ETHZ, EPFL, PSI, EMPA, Eawag and WSL. PHRT’s goals include improving the quality of Personalized Health and Precision Medicine by providing a choice of individual therapeutic strategies for patients. PHRT aims to support and drive transformation in Personalized Health by providing clinicians with access to ETH technologies in order to evaluate their potential contributions to clinical decision making. The PHRT research program complements infrastructure efforts undertaken by the Swiss Personalized Health Network (SPHN) and the Swiss Data Science Center (SDSC). Within the PHRT program, SMOC is an engine for multi-omics data generation, analysis, interpretation. For secure data processing and sharing, SMOC is integrated into the SPHN/BioMEdIT secure computational infrastructure. [http://smoc.ethz.ch](http://smoc.ethz.ch/#overview)

* Clinical Stream: High quality molecular data on the DNA, RNA, Protein, Metabolite and Lipid level for gaining clinical insights.
* Exploratory Research Stream: Integration, visualization and analysis of omics data mapped onto biological networks and pathways.
* SPHN BioMedIT infrastructure integration: For data lineage tracking, secure data management, data sharing, secondary and tertiary analysis.

### Genomics data

From the Clinical Genomic Analysis Center (CGAC), clinical Grade Sequencing (ISO 15189 accredited): Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), RNA Sequencing (RNASeq).

### Proteotyping data

From the Clinical Proteotype Analysis Center (CPAC), Quantitative Proteotyping: Proteotype analysis (DDA, DIA, PRM), Post-translational modification analysis, Spatial proteotype analysis.

### Metabolomics & Lipidomics data

From the Clinical Metabolomics Analysis Center (CMAC): Small molecule analysis, Targeted metabolomics, Untargeted metabolomics, Lipidomics.

## Data analysis

### DNA analysis

We apply best-practices protocols for consistent analysis of genomic data.

For example, the initial data preperation protocols follows the GATK (Genome Analysis Toolkit) best-practices pipeline. This is a widely used and highly recommended method for genomic analysis. The pipeline consists of several steps, including DNA preparation, sequencing, read mapping, variant calling, variant filtering, and functional annotation.

DNA preparation involves the extraction of high-quality DNA from the sample, followed by library preparation and target enrichment. The DNA is then sequenced using a next-generation sequencing (NGS) platform, producing millions of short reads. The GATK analysis begins with read mapping, which involves aligning the short reads to a reference genome. This step is followed by local realignment and base quality score recalibration, which improve the accuracy of the read alignments and base quality scores, respectively.

Variant calling is the next step in the pipeline, which involves identifying differences (variants) between the sample and the reference genome. The GATK variant calling method is based on a probabilistic model, which takes into account the sequencing error rates and other sources of noise in the data. After variant calling, variant filtering is performed to remove low-confidence variants and artifacts. The GATK recommends a set of best-practice filtering criteria, which include quality scores, strand bias, and mapping quality. Finally, functional annotation is performed to determine the potential impact of the variants on protein function and gene regulation. This step involves using publicly available databases and tools to annotate the variants with information such as gene function, conservation, and potential pathogenicity.

Overall, the GATK best-practices pipeline provides a robust and comprehensive approach to genomic analysis, allowing researchers to accurately identify and annotate genetic variants that may be relevant to disease susceptibility, drug response, and other important biological processes.

Once genetic data has been prepared in a consistent and reliable format, interpreation of disease-causing variants can be performed anlong with any statistical analysis protocols. This is done by annotating the data using over 150 databases of all known reliable medical information. To accurately interpret this genetic information we use widely accepted criteria. For example, the ACMG (American College of Medical Genetics and Genomics) criteria for the interpretation of DNA variants provide a framework for the classification of genetic variants based on their potential pathogenicity. The criteria take into account various factors such as the frequency of the variant in the general population, the effect of the variant on protein function, and the inheritance pattern of the associated disease.

The criteria consist of a set of rules and guidelines that are used to assign a variant to one of five categories: pathogenic, likely pathogenic, uncertain significance, likely benign, or benign. Each category is based on a combination of evidence from multiple sources, including population data, functional assays, computational predictions, and familial segregation studies.

This provides a standardized approach for the interpretation of genetic variants that is widely used in clinical and research settings. This helps us to ensure consistency and accuracy in the classification of variants, allowing for better patient care and more efficient discovery of genetic causes of disease.

### Proteotyping analysis

### Metabolomics & Lipidomics analysis

Clinical metabolomics is the study of the small molecule metabolites in biological samples with the aim of understanding their role in health and disease. There are three main methods used in clinical metabolomics analysis: targeted metabolomics, untargeted metabolomics, and lipidomics.

Targeted metabolomics is a targeted and quantitative approach to measuring specific metabolites in a sample. This method uses a pre-defined list of metabolites, which are usually selected based on their biological relevance or prior knowledge. The metabolites are quantified using specific analytical techniques such as liquid chromatography-mass spectrometry (LC-MS) or gas chromatography-mass spectrometry (GC-MS). This method is efficient and cost-effective for large scale studies, as well as for the quantitative analysis of specific metabolic pathways.

Untargeted metabolomics is a comprehensive and exploratory approach to measuring all the metabolites in a sample. This method does not rely on prior knowledge or a pre-defined list of metabolites. Instead, all the metabolites are profiled using high-resolution mass spectrometry techniques, such as LC-MS or GC-MS. This method is more time-consuming and resource-intensive compared to targeted metabolomics, but it provides a more complete picture of the metabolic profile of a sample.

Lipidomics is the study of lipids, which are a class of biomolecules that play important roles in cell membrane structure and function, energy storage, and signaling. Lipidomics involves the identification, quantification, and characterization of the different classes of lipids, such as phospholipids, sphingolipids, and glycerolipids. This method uses similar analytical techniques as in metabolomics, such as LC-MS and GC-MS, but it also involves specific lipidomics platforms and software for lipid analysis and interpretation.

We may rely on commonly used software for each of the methods in clinical metabolomics analysis and small molecule analysis. There are several software platforms available for metabolomics and lipidomics experiments, both targeted and untargeted. For targeted metabolomics, Metabolon’s Analyst and Ideation, Thermo Fisher’s Xcalibur and Proteome Discoverer, and Agilent’s MassHunter and Chemstation are all popular software platforms that provide efficient data analysis, visualization, and interpretation capabilities. For untargeted metabolomics, several open-source software platforms, such as XCMS, CAMERA, and MetAlign in R, Knime, MZmine, and OpenMS, are available for data analysis, visualization, and interpretation. In addition, web-based platforms like MetaboAnalyst and MetPA are also available for untargeted metabolomics analysis. For lipidomics, several software platforms like LIPID MAPS LipidAnalysis, ms-DIAL, LipidXplorer, LipidSearch, and LipidMatch provide similar functionalities for data analysis, visualization, and interpretation of lipidomics data. Overall, these software platforms provide powerful tools for the analysis of metabolomics and lipidomics data, enabling researchers to efficiently process and interpret large amounts of data in a standardized and reproducible manner.