Precision medicine unit

white paper v1.1

Dylan Lawless, PHD[1]

[1]Department of Intensive Care and Neonatology, Universitäts-Kinderspital Zürich, University of Zürich. Dylan.Lawless@uzh.ch

July 17, 2024

Latest version: github.com

Contents

1 Executive summary

The Precision medicine unit at Universitäts-Kinderspital Zürich is proposed to represent the forefront of medical science, using an integrated multi-omics approach to augment paediatric healthcare. This strategy extends beyond genomic data to incorporate a wide array of omics technologies, such as genomics, proteomics, metabolomics, and innovative methods like single-cell sequencing. Together, these technologies can provide a comprehensive understanding of disease mechanisms, which is instrumental in devising personalised treatment strategies for rare and complex paediatric conditions.

Comparable benchmarks have demonstrated clear cost-saving potential through precise diagnostics and targeted therapy. For rare diseases, approximately 40% of probands received a genetic diagnosis [1; 2] and altered critical care management in 77% of diagnosed cases [3]. A well-managed work-flow can result in rapid whole-genome sequencing with a turnaround of 37 hours on average [4]. A forecast, projected into 2030, for the yearly cases of sepsis are shown in **figure 1**. Black and red values show the expected number of deaths with and without precise diagnostics and targeted therapy, respectively.

Establishing these benchmarks as our initial goals is not only technically feasible but also a significant achievement in its own right, setting a foundation for potential further advancements as our understanding and capabilities evolve.

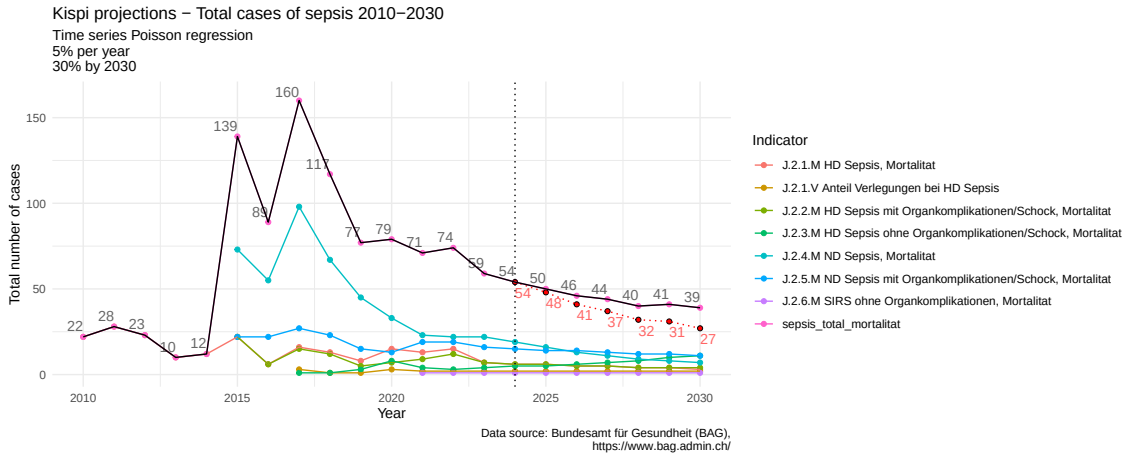

Figure 1: Year deaths due to sepsis at Universitäts-Kinderspital Zürich. This data is based on statistics reported by Bundesamt für Gesundheit (BAG), `https://www.bag.admin.ch/` for years 2010-2022. Time series was performed using Poisson regression to extrapulate the expected outcomes from 2010-2030. Predictions for the cost and number of cases were generated in section 15. DP: Diagnostic procedure. HD: Primary diagnosis. ND: Secondary diagnosis.

2 Introduction

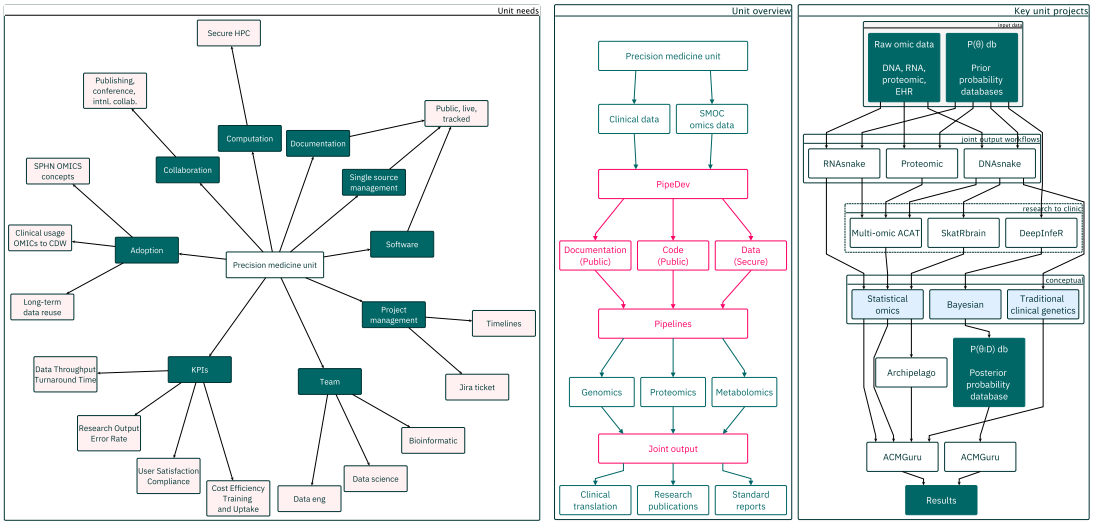

Figure 2: Precision medicine unit overview. The unit needs illustrate the management philosophy for minimal disruption for technical progress. The unit overview illustrates the main flow of information to key products. Key unit products illustrate the flow of information, with technical hurdles avoided using the principles of single-source management and shared open development.

The Precision medicine unit is ready to update our paediatric healthcare by serving an array of advanced capabilities in multi-omics, pioneering research methodologies, purpose-built infrastructure, and rigorous data analysis techniques. Our approach sets a new benchmark for personalised treatment, employing cutting-edge omics services to provide tailored healthcare for the clinic and enrich basic science at Universitäts-Kinderspital Zürich. An overview of our unit is shown in **figure** 2.

Traditionally, medical practices have predominantly adhered to a standard approach, tailoring disease prevention and treatment strategies based on the average response anticipated from a general population. This conventional method has proven effective for certain patients and conditions, but often falls short when individual variances in genetics, environment, and lifestyle are significant factors influencing health outcomes. Precision medicine emerges as a revolutionary paradigm, diverging from the one-size-fits-all approach to embrace these individual differences. This innovative medical strategy integrates detailed layers of genomic, environmental, and lifestyle data to tailor personalized healthcare solutions. The evolution of precision medicine has been markedly propelled by significant advances in biomedical research, affecting millions worldwide. It aims not only to refine disease treatment and prevention but also promises enhanced outcomes by aligning more closely with each person's unique biological makeup. This approach has profound implications across various domains of healthcare, including oncology, pharmacogenomics, and the management of rare diseases, setting a new standard in personalized healthcare. Our partners are implementing precision medicine approaches at the Swiss Institute of Bioinformatics (SIB), The LOOP Zurich, CHUV, and all private industry leaders such as Roche, Genentech, and Pfizer.

There have also been many failed promises where precision medicine did not deliver. However, these failures are typically due to the logistical difficulties of creating an efficient system. This is not due to inherent problems with the technology. Success has been demonstrated by a number of examples in the UK, US, and UAE, which resulted from careful planning and development.

The Precision medicine unit at Universitäts-Kinderspital Zürich can be ready to collaborate closely with the Health 2030 Genome Center, part of the Swiss Multi-Omics Center (SMOC) (http://smoc.ethz.ch), to harness state-of-the-art genomic services. This partnership underpins our cutting-edge approach in pediatric healthcare, enabling us to deliver personalised treatment strategies through comprehensive multi-omic data analysis. As a collaborative effort among major Swiss universities and university hospitals, including UNIBE, Inselspital, UNIL, CHUV, UNIGE, HUG, and EPFL, the center exemplifies a comprehensive approach to personalized health care and genomic research. SMOC integrates expertise across genomics, transcriptomics, proteomics, and metabolomics. This integrated multi-omic analysis is essential for the thorough exploration and understanding of clinical specimens. Accredited under ISO 15189, the center ensures the highest standards in sequencing and data analysis, leveraging state-of-the-art technologies such as the Illumina Novaseq6000 platforms and TruSeq DNA PCR-Free library preparation.

We will employ essential services such as:

- Clinical-grade whole genome sequencing (WGS)
- Clinical-grade whole exome sequencing (WES)
- Clinical-grade RNA sequencing (RNA-seq)
- Fast turnaround time for clinical-grade WGS, WES, and RNA-seq
- Low-coverage WGS with variant imputation
- Viral pathogen WGS
- Microbiome WGS

Our subsequent analysis occurs at the cutting edge of technology, employing the most reliable reference genomes based on GRCh38 and adhering to stringent guidelines for handling sensitive clinical data, as outlined by the SPHN/BioMedIT network on the sciCORE platform (https://sphn.ch/network/projects/biomedit/). This framework supports the SwissPedHealth initiatives, enhancing the capacity for robust clinical diagnosis and research. We are also thus preparing to meet the rapidly approaching national expectations. These expectations are demonstrated by projects like the Swiss Federated Genomics Network (SFGN) and Genome of Switzerland (GoS) which aim to sequence clinical-grade genomes on a national scale, contributing to European genomic initiatives and supporting the advancement of genomic medicine across the region.

By capitalising on these integrated expert services we can provide analysis within the Precision medicine unit:

- Sequencing data processing
- Genome variant calling
- Variant clinical annotation and prioritising
- RNA expression profiling
- Single cell RNA-seq
- Multi-omic joint analysis of DNA, RNA, proteomics

The Precision medicine unit will therefore be able to provide clinical-grade results with fast turnaround time - from DNA to variant interpretation and functional effect. Material will be shipped from Universitäts-Kinderspital Zürich to SMOC where high-throughput OMICs performed. OMIC analysis will then be performed on our secure high-performance computer infrastructure. Clinical reports containing actionable results can be provided to the clinic. Rich datasets will be generated for research and discovery.

3 Key products

Products of the Precision medicine unit include ACMGuru (v1.0) for variant classification and interpretation, DeepInfeR (v1.0) for defining the posterior probability of every genetic determinant of disease based on all known public data, Archipelago (v1.0) for statistical interpretation of variant set association testing (VSAT), SkatRbrain (v0.2) for automated statistical genomics, multi-omic ACAT (v0.1) for multi-omic joint analysis of VSAT by aggregated Cauchy association test (ACAT), DNAsnake (v0.1) for DNA pre-processing RNAsnake (v0.1) for RNA pre-processing, and the documentation of pipelines in Pipe-Dev docs DNAsnake (v0.1).

3.1 Product example: advanced DNA sequencing data preprocessing

DNAsnake (v0.1) represents a vital product from our Precision medicine unit, meticulously designed to preprocess WGS DNA data for use in clinical genetics reporting, statistical analysis, and machine learning applications. Employing the Genome Analysis Toolkit (GATK), DNAsnake (v0.1) standardises the preparation of DNA sequencing data to ensure consistency and reliability across diverse analytical applications.

DNAsnake (v0.1) is engineered around the GATK best practices for DNA sequence data preprocessing. This workflow is integral to producing high-quality, clinical-grade DNA data outputs, which are crucial for downstream processes like variant interpretation in ACMGuru and assessing genetic determinants of disease with deepInfeR.

The DNAsnake (v0.1) workflow is detailed and robust, encompassing several critical stages of DNA preprocessing:

1. **Quality Control and Pre-processing**: Initial receipt of raw FASTQ files followed by quality control assessments using tools like FastQC and subsequent trimming with Trimmomatic.
2. **Alignment**: Alignment of sequences to the GRCh38 reference genome using the Burrows-Wheeler Aligner (BWA).
3. **Post-alignment Optimization**: Includes marking duplicates with Picard Tools, and base quality score recalibration (BQSR) with GATK's BaseRecalibrator and ApplyBQSR tools.
4. **Variant Calling**: Utilizing GATK's HaplotypeCaller for calling germline SNPs and indels, followed by variant quality score recalibration (VQSR) to ensure high-quality variant calls.
5. **Output Generation**: Production of annotated, processed BAM and VCF files ready for comprehensive genetic analysis.

Processed outputs from DNAsnake feed directly into:

- ACMGuru (v1.0), for detailed variant classification and interpretation in line with ACMG guidelines.
- DeepInfeR (v1.0), which utilizes the processed data to calculate the posterior probabilities of genetic variants, influencing disease phenotypes based on extensive public data repositories.

By automating the WGS DNA data preprocessing with DNAsnake (v0.1), our unit will achieve:

- **Standardisation and reproducibility**: Ensures that all samples are processed through a uniform pipeline, reducing variability and enhancing the reliability of results.
- **Efficiency and scalability**: Capable of handling large-scale datasets with the flexibility to accommodate increasing data volumes without sacrificing performance.
- **Integration and Interoperability**: Seamlessly interfaces with other analytical tools and databases, promoting a cohesive and integrated approach to precision medicine.

DNAsnake (v0.1) exemplifies our commitment to delivering state-of-the-art solutions for genetic data preprocessing. As an element of our single-source management strategy, it not only supports but enhances the capabilities of our precision medicine initiatives, ensuring that data used across various platforms is of the highest quality and utility.

4 Multi-omics integration

Our unit will use whole genome sequencing (WGS) to identify both common and rare genetic variants, thereby establishing a foundational genetic blueprint of disease predisposition. In complement to WGS, RNA sequencing (RNAseq) provides insights into gene expression changes across different physiological states or in response to treatments, shedding light on the functional consequences of genetic alterations.

Furthermore, the incorporation of proteomics and metabolomics technologies offers a detailed examination of the proteome and metabolome, which are in closer correspondence to the phenotype. These methodologies facilitate the monitoring of biochemical activities and protein functions that have direct impacts on disease phenotypes, effectively bridging genotypic information with tangible biological and clinical outcomes.

In addition, routine clinical measurements and electronic health records (EHRs) will be integrated with omics data, augmenting the granularity and practicality of our findings. This synthesis enables tailored, real-time therapeutic interventions and continuous monitoring of disease progression, markedly enhancing the precision of diagnostics and personalised treatment plans.

5 Project management

We will manage intricate projects that encompass multiple stakeholders and complex data streams. Employing Jira, an advanced project management tool, we can clearly oversee tasks, timelines, and dependencies, ensuring that every phase of our projects conforms to our exacting standards and strategic objectives. This dynamic approach underpins our adaptability, facilitating the efficient execution of our research and clinical activities.

The documentation and outputs of our research will be rigorously curated from the first step to ensure transparency and accessibility, adhering strictly to our existing and future standards. Sensitive data will be secured and protected in BioMedIT to adhere with patient confidentiality and data integrity. Data transfers between BioMedIT and Kispi will follow standard guidelines and use secure methods such as sFTP backed by data transfer agreements. Stringent data stewardship is critical for maintaining trust and ethical standards in our research.

6 Data flow

The data generation and analysis process involves several stages, beginning with sample collection and culminating in the presentation of final results. This process is designed to integrate with existing clinical databases using newly developed OMICs concepts based on the SPHN ontology. The process is summarised in **figure 3.**

We have worked with SPHN and TheHyve to develop new concepts which cover the generation of sequencing data and analysis, recently published in van der Horst et al. [5] as illustrated in **figure 4**. We have also developed the concepts which cover the final outputs of downstream analysis results for omic results. This work enhances the integration of omics data into the SPHN Semantic Interoperability Framework, which primarily handles clinical routine data.

Our new genomics extension enriches the framework to include comprehensive descriptions of genomics experiments, encompassing both clinical and research applications. It outlines the entire omics process flow, detailing steps from sample processing to data analysis, including specifics like library preparation and sequencing analysis. The extension also integrates additional omics metadata, such as details on sequencing instruments and quality control metrics. By aligning with established semantic models and leveraging common biomedical vocabularies (e.g., EDAM, OBI, and FAIR genomes), it promotes semantic interoperability and aims to FAIRify data for shared use within the Swiss network, enhancing data reuse in a unified knowledge graph.

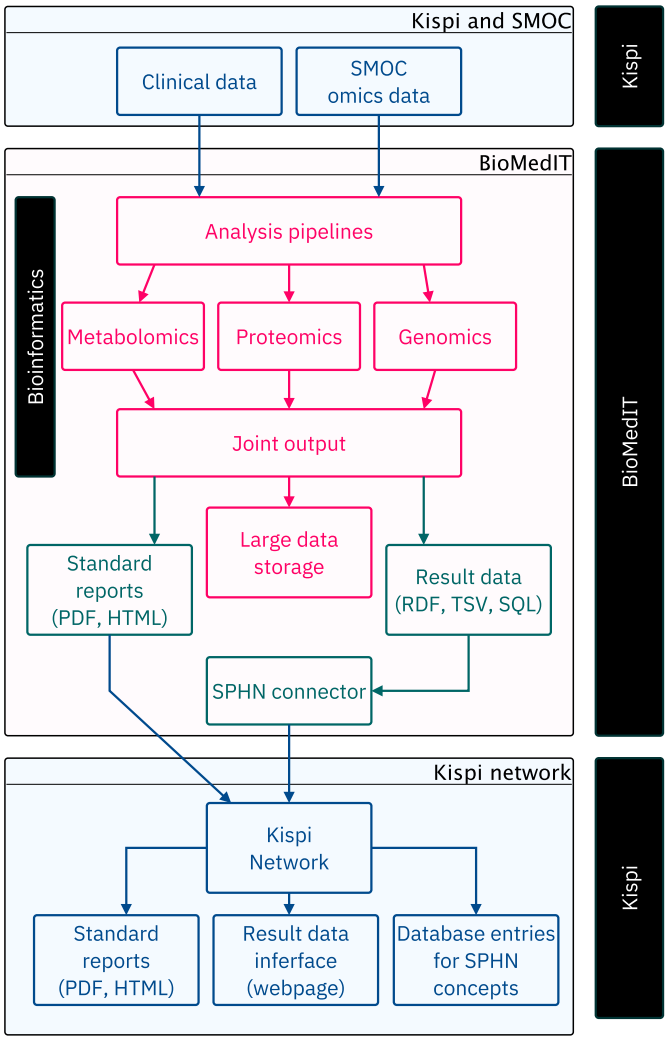

Figure 3: Precision medicine unit data flow from sample collection to final result presentation. **Sample collection**: Sample collection occurs in Universitäts-Kinderspital Zürich. SMOC typically processes most physical samples of DNA, RNA, serum, or other tissues. Multi-omic data is generated and transfered to BioMedIT. **Analysis on BioMedIT**: Bioinformatic analysis pipelines process the data and produce a main analysis output which is stored long-term. Key actionable results from this large dataset are prepared according to SPHN ontology, using concepts such as “sequencing assay” and “omic result”. **Transfer results to Universitäts-Kinderspital Zürich**: Two datasets are prepared before transfer to the Universitäts-Kinderspital Zürich network. (1) The data for which we have SPHN concepts and is suitable for a clinical data warehouse is prepared in TSV, SQL, RDF, or other format. (2) Supplemental reports with extensive metadata, visualisations, and contextual information in formats such as TSV, PDF, HTML. Both datasets are transfered to the Universitäts-Kinderspital Zürich network for database integration and file storage, respectively.

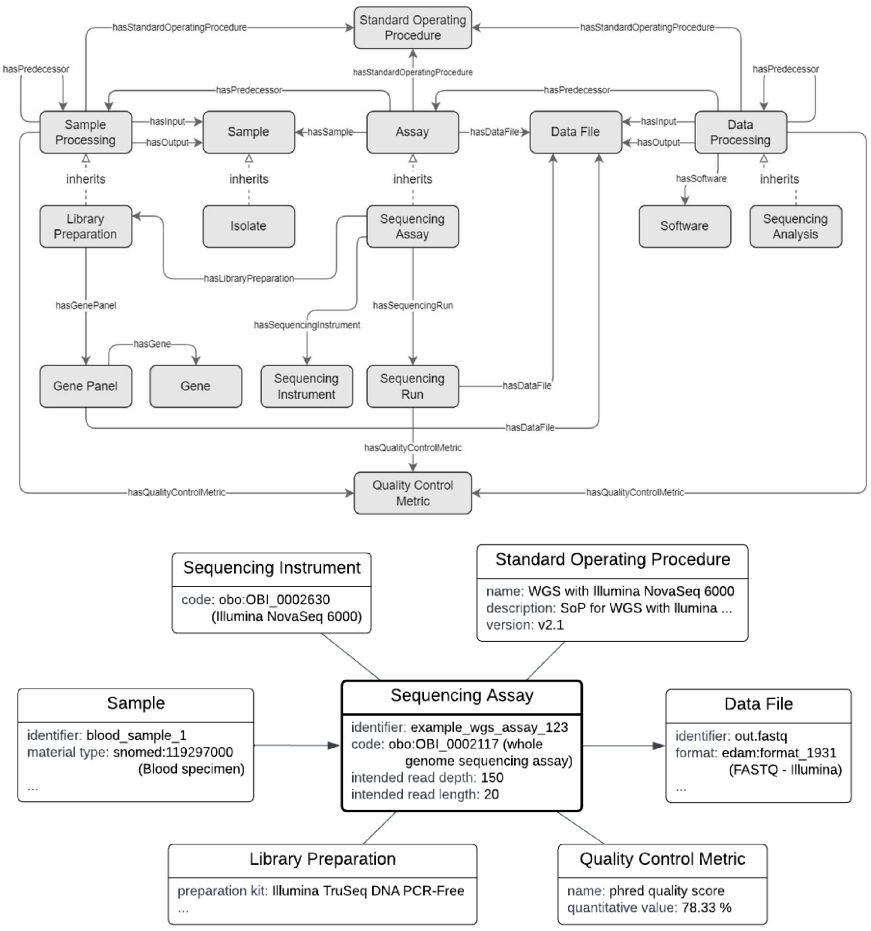

Figure 4: Figure extract and text quoted from van der Horst et al. [5]: (A) Basic excerpt of the schema for the (gen)omics process flow. (B) Diagram visualising an instance of a sequencing assay that analyses one sample and produces one FASTQ file. “Sequencing Assay” concept, together with its “Instrument”, “Library Preparation”, “Standard Operating Procedure”, and “Quality Control Metric” concepts from which it is composed.

6.1 Sample collection and initial processing

- Location: Universitäts-Kinderspital Zürich.
- **Details:** Sample collection encompasses various biological materials including DNA, RNA, serum, and other tissue types. These are initially processed by the Swiss Multi-Omics Center (SMOC), which is responsible for the physical handling and preliminary omics data generation.

6.2 Data transfer and bioinformatics analysis

- **Transfer:** The raw multi-omic data generated by SMOC is transferred to BioMedIT using secure protocols such as sFTP.
- **Bioinformatics processing:** At BioMedIT, advanced bioinformatics pipelines are employed to analyse the data. This includes comprehensive analyses across metabolomics, proteomics, and genomics disciplines.
- **Outputs:** The main outputs from these analyses include:
  - A large dataset stored for long-term reuse and research purposes.
  - Standard reports generated in formats such as TSV, PDF, and HTML.
  - Result data formatted in RDF, SQL, TSV, which are then adapted to meet the SPHN connector's requirements for merging into clinical data warehouses.

6.3 Data conversion and integration

- **Conversion:** Key actionable results are extracted from the large dataset and prepared according to our reporting evidence guidelines and formatted using the SPHN ontology. This preparation uses specific OMICS concepts such as "omic result" to ensure that the data can be seamlessly integrated and interpreted within the clinical framework.
- **Integration:** The processed results are converted to fit the database requirements of the hospital's clinical data warehouse.

6.4 Presentation of final results

- **Internal Network Transfer:** Outputs, including the standard reports and result data, are transferred back to the Universitäts-Kinderspital Zürich network. This step is crucial as clinicians do not have direct access to secured BioMedIT servers.
- **Access and Presentation:** Final analysis results are made accessible to clinicians through an internal webpage and downloadable TSV, PDF, or other formats. This ensures that the results are readily available for clinical decision-making and further research. The clinical data warehouse will maintain the main omic result data in RDF, SQL, or other suitable formats that best match the current system.

By aligning with the SPHN RDF ontology and implementing it through newly developed OMICS concepts, this data flow ensures that genomic and other omic data types are integrated into the hospital's clinical operations, enhancing the capacity for precision medicine and personalised patient care. The entire process is illustrated in **figure** 3, providing a visual representation of the data flow from sample collection to final result presentation within the hospital's infrastructure.

7 Clinical implementation and future prospects

Our multi-omics data inform precision medicine protocols, guiding treatment plans that are specifically tailored to individual genetic profiles, thus enhancing treatment efficacy and minimising adverse effects. The prognostic models developed from our multi-omics data facilitate early interventions and enhance disease management strategies.

Looking to the future, we are dedicated to integrating single-cell sequencing and additional omic technologies, such as lipidomics and glycomics. These advancements are expected to considerably refine our understanding of diseases at a resolution that was previously unattainable. This development is anticipated to improve our diagnostic capabilities and therapeutic interventions, ensuring that Kinderspital Zurich continues to lead in paediatric healthcare innovation.

8 Strategic collaborations

Collaboration forms a cornerstone of our operational philosophy. We have partnerships both nationally and internationally, enhancing the capabilities and influence in paediatric precision medicine. These collaborations will extend beyond academic institutions; we will actively engage with healthcare providers, policymakers, and the pharmaceutical industry to ensure that our research has a wide-reaching and significant impact.

By participating in global research consortia and contributing to international databases, we share our findings and innovations, fostering a culture of open science and continuous improvement. These strategic partnerships amplify our research capabilities and enable us to remain at the forefront of technological and methodological advancements in precision medicine.

This proactive approach to project management and strategic collaboration positions by the Precision medicine unit will help use to become leaders in translating scientific research into clinical outcomes, significantly advancing the field of paediatric healthcare.

9 Funding model and sustainability

The Precision medicine unit will benefit from a diverse funding model that includes institutional support, competitive research grants, and strategic partnerships with industry leaders. This multifaceted approach provides a robust financial foundation that helps ongoing research and development in paediatric precision medicine. Institutional support from Kispi and its affiliated academic entities should furnish a stable base of funding that sustains core facilities and staffing requirements.

Beyond traditional funding avenues, we may proactively establishes collaborations with pharmaceutical companies and biotechnology firms. These partnerships frequently catalyse joint research projects and grant access to avant-garde technologies and additional streams of funding. By harmonising our research objectives with the needs of our industry partners, we not only secure essential resources but also ensure that our research outputs are pragmatically applicable and rapidly translatable into clinical settings.

Furthermore, we are dedicated to integrating sustainable practices to sustain all facets of our research activities, from data management and analysis to resource utilisation. Our goal is to build a system that is redundant to decay. To do so, we must produce tools that are sensible for newcomers to learn, provide a consistent method of production and usage so that they can be updated and maintained without dependency on single members.

The most common research funding pathways for academic members of the Precision medicine unit are shown in figure 5 and links to funding opportunities that are also supported by UZH assistance are provided in table 1.

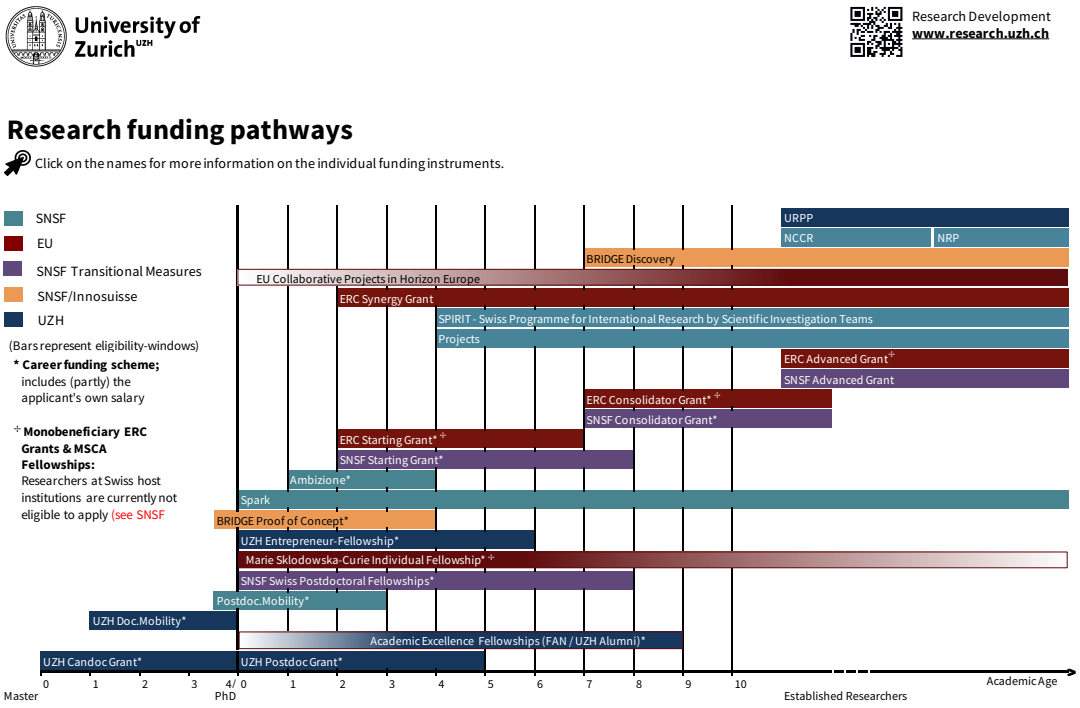

Figure 5: Most common research funding pathways for academic members.

Program	
Alumni-Fonds, UZH	PRIMA (Promoting Women in Academia, SNSF)
Ambizione (SNSF)	Privatdozenten-Stiftung, UZH
AXA Research Fund	Research Equipment (R'Equip, SNSF)
Bridge: Discovery	Scientific Exchanges (SNSF)
Bridge: Proof of Concept	SNIS Project Funding
Centers of Competence	SNSF Advanced Grant
Citizen Science Seed Grants (UZH)	SNSF Consolidator Grant
Collegium Helveticum Early-Career Fellowships	SNSF Eccellenaz Professorial Fellowship
Collegium Helveticum Senior Fellowships	SNSF Project Funding
Competitive Sabbaticals, UZH	SNSF Starting Grant
Cotutelle de Thèse	SNSF Swiss Postdoctoral Fellowship
DIZH Innovation Program	Spark (SNSF)
DSI Infrastructure & Lab Call	SPHN (Swiss Personalized Health Network)
ERC Advanced Grant	SPIRIT (SNSF)
ERC Grants	Swiss 3R Competence Centre (3RCC)
FAN Academic Excellence Fellowships, UZH	Excellence Scholarships for Foreign Scholars
Foreign Government Scholarships	Research and Innovation Cooperation Programs
Foundation for Research in Science and the Humanities at UZH	Swiss Personalized Health Network (SPHN)
GRC Grants (UZH)	Swiss-European Mobility Programme
GRC Travel Grant (UZH)	swissuniversities: Project Contributions
GSPI Impact Collaboration Programme	TRANSFORM (UZH)
Heidi Ras	University Research Priority Programs (URPP)
Horizon Europe	UZH Candoc Grant
Innosuisse Projects	UZH Doc.Mobility
Marie Skłodowska-Curie Postdoctoral Fellowships	UZH Entrepreneur Fellowships
National Centers of Competence in Research (NCCR)	UZH Global Strategy and Partnerships Funding Scheme
National Research Programmes (NRP)	UZH Index of Foundations (German only)
NIH and other US Grants	UZH Postdoc Grant
Postdoc.Mobility (SNSF)	

Table 1: List of funding opportunities which include organisational support by UZH. (Note hyper-links in electronic PDF version.)

10 Key performance indicators

To ensure that our research and clinical implementations are both effective and efficient, the Precision medicine unit will track performance using a well-defined set of Key Performance Indicators (KPIs). These KPIs are carefully chosen to reflect crucial aspects of our operations, from data management to user engagement and compliance. We also aim to take the production burden off individuals and instead place it on the joint group effort to determine our own pace, remaining flexible to for worker satisfaction and development.

Our KPI framework includes the following metrics, each designed to provide insights into specific operational aspects:

- **Data Throughput:** Measured by the total gigabytes (GB) of data processed weekly, this indicator helps assess our capacity to handle large-scale genomic datasets.
- **Turnaround Time:** This KPI tracks the duration from sample receipt to the delivery of the report, crucial for evaluating our efficiency in processing and reporting.
- **Research Output:** Annually quantified by the number of papers published and tools developed, reflecting our contribution to scientific knowledge and tool innovation.
- **Error Rate:** This is calculated as the number of errors per 100 analyses, providing a clear measure of the precision and reliability of our data analyses.
- **User Satisfaction:** Determined by averaging scores from user surveys, this metric gauges the satisfaction of clinical staff and researchers with our tools and reports.
- **Compliance:** Measured by the percentage of successful compliance checks, this KPI ensures that our operations adhere to all relevant regulatory and ethical standards.
- **Cost Efficiency:** This is evaluated by the cost per analysis or report, helping us monitor financial efficiency in our operations.
- **Training and Uptake:** The percentage of cases using new tools measures the adoption rate of our latest methodologies and technologies, indicating the effectiveness of our training programs.

11 Single-source management in precision medicine

In the fast-evolving field of precision medicine, managing documentation, data, and software development requires a robust approach to maintain consistency and accuracy across various studies, clinical trials, and software platforms. Our group adopts a **single-source management** philosophy, centralising shared components to reduce redundancy and prevent discrepancies. This idea fit's with the "Everything as code" philosophy.

11.1 Principles of single-source management

Single-source management is predicated on the idea that all derivative documents and systems should draw from a common repository of information, ensuring that changes in one area are automatically propagated to all related areas. This approach is particularly advantageous in environments where:

- Documentation needs to be consistent across multiple reports and publications.

- Software developments are closely tied to ongoing clinical trials and studies.
- Data from these trials and studies are frequently updated and must remain synchronised across platforms.

Figure 6 illustrates why this is important. If a user requires a piece of critical information (for example, the number of patients in a study), it must be checked and copied. If that information later updates then all of the manual steps to find that information must also be repeated. When scaled across multiple different projects/uses, it rapidly degrades productive time into “busy work” for the user as well as leading to numerous critical errors. Figure 6 illustrates why this is important. This illustration shows a two simple examples of where single-source management is used for two and six use cases, respectively.

If, for example, the contents of “raw data” is updated as a project updates, all of the subsequent uses of this information can be automatically updated in a single step rather than repeating work numerous times. !! N.B. - Every edge which connects two nodes represents a task which must be performed; either manually or automatically using single-source management.

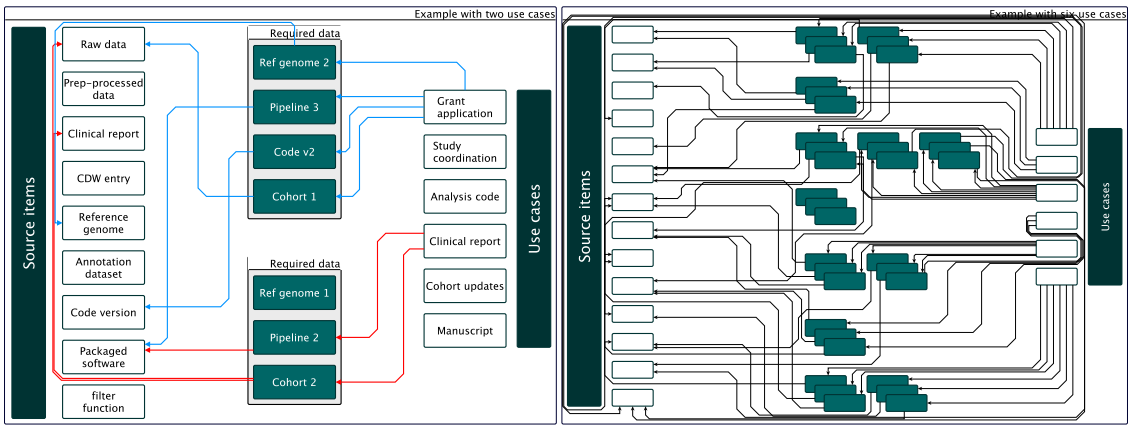

Figure 6: Single-source management examples within the Precision medicine unit. As the number of uses for information increases, the scale of tracking grows massively. Every edge which connects two nodes represents a task which must be performed. Without automation, manually performing each task causes a bottleneck that can halt progress. Single-source management allows us to update every edge automatically.

11.2 Implementation example in our documentation

Our LaTeX documents exemplify this single-source management system. The document that you are reading has been generated using this system. For example, all of the highlighted key texts used in this document are sourced from a central repository so that they are always up-to-date. Key elements such as common variables, formatting settings, and version-controlled content are centralised. We maintain a structured directory layout for efficient resource management:

```
/resources
    /head.tex           % Standardized layout and color definitions
    /variables          % Common variables across projects
        /variables.tex  % Definitions of constants, pipeline names, etc.
```

```
/main_document_1.tex    % Specific document drawing from the resources
/main_document_2.tex    % Another document sharing the same resources
```

The same approach is taken in our software development, analysis protocols, data-flow, etc. There is no philosophical separation between any type of information - a document name is treated with the same respect as raw data from a patient. There is a single source of truth for information and all subsequent uses of that information directly source the origin, ideally including tracked changes. Numerous sources of variables can be used, including the database repository, document layouts, and other configurations. The variables source contains declarations for key terms and concepts used across various documents, highlighted using a specific color to indicate their importance and frequent usage. These variables serve not only as identifiers within texts/code/etc. but also help to connect the reader/user with recurring themes and important concepts across multiple documents.

11.3 Benefits

The benefits of single-source management in our precision medicine research are manifold:

- **Increased efficiency**: Updates need only be made in one place, automatically propagating to all documents and systems.
- **Enhanced accuracy**: Reduces the risk of discrepancies between documents, as all draw from a unified source.
- **Improved collaboration**: Team members can work simultaneously on different documents, confident that they are using the most current and consistent information.

Implementing this system requires rigorous discipline and a well-organised file management structure, but the return on investment in terms of time saved and error reduction is substantial.

12 IVDR compliance documentation and version controls

12.1 Introduction to IVDR

The In Vitro Diagnostic Medical Devices Regulation (IVDR) is an essential legislative framework that governs the safety and performance of in vitro diagnostic medical devices (IVDs) within the European Union. Implemented to enhance patient safety and ensure high standards of quality, the IVDR was adopted alongside the Medical Devices Regulation (MDR) to update and replace directives established in the 1990s, reflecting significant technological and scientific advancements in the sector.

The IVDR, Regulation (EU) 2017/746, was published in May 2017 and became fully applicable on May 26, 2022. This regulation introduces a risk-based classification system, stricter pre-market scrutiny, and enhanced transparency through a comprehensive EU database on medical devices (EUDAMED). It mandates clear obligations for economic operators, including manufacturers, importers, and distributors, and strengthens post-market surveillance and vigilance requirements. This regulatory framework is crucial for ensuring that in vitro diagnostic devices meet the latest standards for safety and efficacy, supporting

the health and well-being of Swiss and EU citizens. For detailed guidance and documents endorsed by the Medical Device Coordination Group (MDCG), we refer to the European Commission's dedicated page on IVDR: https://health.ec.europa.eu/medical-devices-sector/new-regulations/guidance-mdcg-endorsed-documents-and-other-guidance_en.

The relevant resources for detailed guidelines can be found here:

- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Quality Guidelines: https://www.ich.org/page/quality-guidelines
- European Medicines Agency (EMA) Scientific Guidelines: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines
- Eudralex - The rules governing medicinal products in the European Union: https://health.ec.europa.eu/medicinal-products/eudralex_en

We are developing a unique method of single-source management combined with an everything-as-code approach to systematically address IVDR compliance from the outset. In many commercial companies, regulatory compliance is typically managed by a dedicated team. However, by adopting our planned system of linked documentation from the start, we can automate much of this process. This proactive approach makes the entire audit process more straightforward and less stressful, provided our members are on board with using this system.

12.2 Example: Compliance and audit

During audits, it is common for auditors to select a specific piece of information, such as a patient ID, a reference file, or a code repository, and then verify compliance by reviewing all related documentation and data. Our system simplifies this process by automatically linking and tagging every relevant document and dataset to the selected item.

To perform an audit, we execute a command that initiates a thorough search through our integrated system, pulling together all connected data and documentation. This network of information, pre-prepared and linked, serves as the foundation for our audit documentation. The audit reports are dynamically populated with the relevant variables pulled from our single-source variable files, which meticulously list every piece of reference data and code repositories, including metadata. This ensures that each piece of information is traceable and verifiable at any moment.

This approach not only secures ongoing compliance but also enables us to demonstrate our adherence to IVDR standards effectively and transparently at any time. The efficiency of this process significantly simplifies the audit procedure, reducing the workload typically managed by a dedicated regulatory team and thereby enhancing our operational efficiency and compliance reliability.

The following example demonstrates some of the variables used to automatically retrieve the linked data.

12.2.1 Document Control

- Document ID: DS-IVDR-003

- Version: 1.2
- Approval date: [Insert Date]
- Review cycle: Annually or upon significant changes to the pipeline, software, or regulatory guidelines.

12.2.2 Introduction

- Purpose: To detail the version control practices and database management strategies employed in DNAsnake (v0.1) for IVDR compliance.
- Scope: Includes software version control, database management, and change tracking for compliance with regulatory requirements.

12.2.3 System overview

- Pipeline Name: DNAsnake (v0.1)
- Version Control System: Git
- Components:
  - GATK for variant detection.
  - Ensembl VEP for annotation.
  - GRCh38.p13 as the reference genome.

12.2.4 Version control management

Git repository details:

- DNAsnake (v0.1) Repository: `https://github.com/SwissPedHealth-PipelineDev/docs`
- DNAsnake (v0.1) commit: 581d0ed0f67ad86669ffcb2d2a03f22638fda1ae
- GATK repository: `https://github.com/broadinstitute/gatk`
- GATK commit: 59c9c1bba1c3edf6624468fd4f81f4fa2fe3fbae
- Ensembl VEP repository: `https://github.com/Ensembl/VEP_plugins`
- Ensembl VEP commit: 74c9315623660e622effb7572c1ed21e6700c2ea
- Example commit check: Use `git log --oneline` to review recent commits and ensure updates are tracked.

Release management:

- Release tag used: `git tag -a v0.1-m "Release version 1.0 for clinical deployment"`
- Check out release: `git checkout tags/v0.1`

12.2.5 Reference genome details

- Internal usage documentation: https://swisspedhealth-pipelinedev.github.io/docs/pages/ref.html
- Genome name: GRCh38.p13
- Genome version: GCA_000001405.15_GRCh38_no_alt_analysis_set

- Source and location: `ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz`
- Checksum/Hash Value: To verify integrity before use.
- Updating reference data: Process for updating reference genome versions documented in `REF_UPDATE.md` at the repository root.

Note: see how PacBio does this:`https://github.com/PacificBiosciences/reference_genomes/blob/main/reference_genomes/human_GRCh38_no_alt_analysis_set/human_GRCh38_no_alt_analysis_set.sh`

12.2.6 Metadata and compliance tracking

Metadata repository integration:

- System name: Clinical metadata repository (CMR)
- Metadata includes: Version tracking, processing steps, QC metrics, and usage logs.
- Automated metadata logging: Scripts to pull version info and log it in CMR.
- Example metadata retrieval command: `git describe --tags --always`

12.2.7 Compliance tagging and audit trails

Audit trail setup:

- Log commands: `git log --since="YYYY-MM-DD" --grep="IVDR_Compliant "`
- Audit reports: Automatically generated from CMR data, including Git tag and commit references for audit periods.

12.2.8 Case study and example documentation

- Example usage: Document a hypothetical use case where DNAsnake (v0.1) processes a sample, including Git commit IDs for software and reference data used.
- Compliance Example: Show a compliance check with links to Git commits proving usage of the validated versions.

12.2.9 Conclusion

Summary: Emphasises the robustness of the version-controlled environment in maintaining IVDR compliance and ensuring the reproducibility and reliability of clinical diagnostics.

12.2.10 Appendices

- A. Git command reference: Common commands used for version control in the project.

- **B. Change management logs:** Example format for documenting significant changes.
- **C. Compliance tag definitions:** Detailed explanation of tags used for IVDR compliance.

13 Future directions and innovation

As we look to the future, the Precision medicine unit is set to incorporate more advanced technologies, such as artificial intelligence (AI) and machine learning algorithms, into our workflow. These technologies are expected to substantially enhance our capability to analyse and interpret the copious amounts of data we generate, thereby facilitating the development of more accurate diagnostic tools and more efficacious therapeutic strategies.

In parallel with technological expansion, we are planning to initiate new collaborative projects that utilise our multi-omics platforms to delve into previously unexplored areas of paediatric health and disease. These projects will strive to identify novel biomarkers, elucidate complex disease mechanisms, and forge innovative treatments tailored to the genetic and molecular profiles of individual patients.

Our ongoing commitment to innovation and excellence ensures that we will continue to enhance health outcomes for children and further afield, significantly shaping the future of paediatric medicine with each discovery we make.

14 Future products

- Noninvasive prenatal testing using circulating DNA and RNA [6].
- Pharmacogenomics drug dosages.
- Pharmacogenomics drug interactions.
- Pharmacogenomics cost effectiveness.
- DeepInfeR

14.1 Pharmacogenomics at Kispi

Pharmacogenomics has the power to find direct effects of personal genetics on the metabolism and effectiveness of drugs used in the clinic. **Figure 7** illustrates an example of variant effect from gene-to-protein and how some gene-drug interactions are identified. Our pharmacogenomics strategy has a tripartite approach which is designed to personalise and enhance treatment efficacy while managing healthcare expenditures effectively. Pirmohamed [7] has reviewed the current status and future perspectives of such approaches.

Firstly, the pharmacogenomics of drug dosages focuses on adjusting medication based on genetic profiles to optimise therapeutic effects and minimise adverse reactions [8]. By analysing a patient's genetic markers, the Precision medicine unit can predict metabolic rates and drug absorption, ensuring that each child receives a dosage that maximises benefit while reducing the risk of toxicity. This tailored approach not only promises better health outcomes but also reduces the trial and error typically associated with standard dosing practices.

Secondly, understanding pharmacogenomic drug interactions is crucial in a paediatric setting where polypharmacy can be common, especially in complex cases such as cancer or chronic diseases. The **Precision medicine unit** leverages genetic insights to foresee and mitigate adverse drug-drug interactions. This is achieved by mapping genetic variants that affect drug metabolism enzymes, transporters, and receptors, thereby providing a clear picture of potential interactions before they can cause harm.

Lastly, the aspect of cost-effectiveness in pharmacogenomics cannot be overstated. By integrating pharmacogenomic data into clinical decision-making, the **Precision medicine unit** aims to reduce not only the financial burden associated with ineffective treatment but also the costs related to adverse drug reactions. This strategic incorporation of pharmacogenomics is expected to lead to more efficient use of healthcare resources, shorter hospital stays, and fewer unnecessary tests and procedures [9].

The Cytochrome P450 (CYP) enzymes are an especially useful target of pharmacogenomics. These are crucial for the metabolism of many xenobiotics and endogenous compounds, making them key players in drug metabolism. Variations in the 15 encoding genes significantly influence drug interactions and efficacy. The Pharmacogene Variation (PharmVar) consortium maintains a repository that catalogues these genetic variations, aiming to standardize nomenclature across the pharmacogenetics community [10]. This facilitates understanding and predicting drug metabolism, disposition, and response. Other essential resources in this field include the Pharmacogenomic KnowledgeBase, DrugBank, and the Clinical Pharmacogenetic Implementation Consortium, which further support pharmacogenomic research and clinical implementation.

Over the last six years (2019-2024), our methods have been used for teaching in the EPFL master's degree coursework and hands-on projects in BIO491 - MSc New tools & research strategies in personalized health and tutorials.

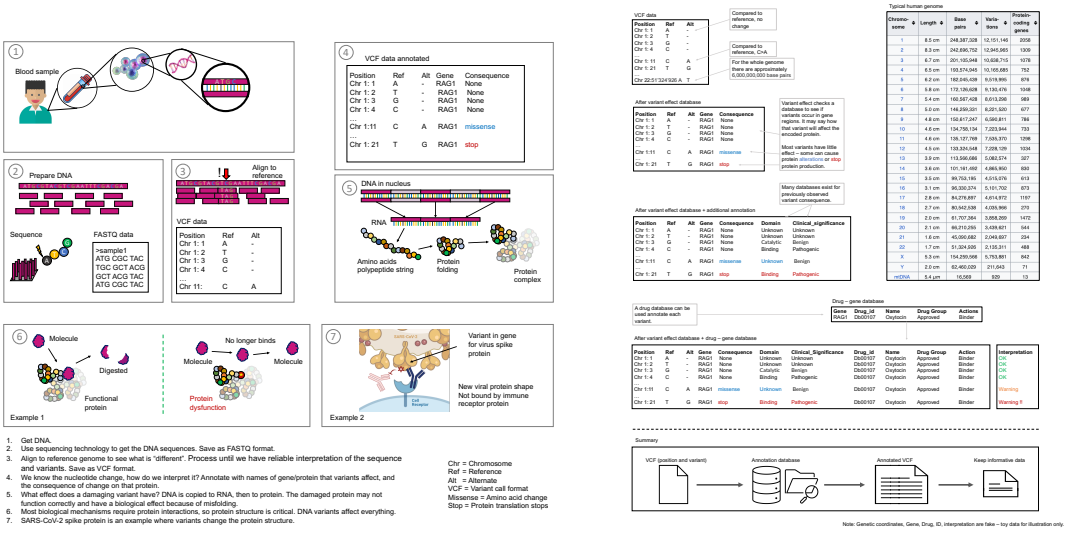

Figure 7: **Precision medicine unit** pharmacogenomics. Known genetic variants produce measurable effects on protein and function. By applying the pharmacogeomic knowledge base to patient's personal genomic data, we can determine the gene-drug effect. This allows us to provide information about drug metabolism, dosage, indication, and cost effectiveness.

14.2 DeepInfeR

A significant development within the Precision medicine unit is the DeepInfeR (v1.0) initiative, which embodies our forward-thinking approach in genetic research. This project aims to pre-calculate the probability of each genetic variant, whether known or novel, in causing any given set of diseases using a comprehensive Bayesian framework. This initiative sets a new benchmark in predictive medicine by integrating extensive prior knowledge and observational data into a cohesive statistical model.

The methodology behind DeepInfeR (v1.0) involves several rigorous steps:

1. **Variant Data Collection and Annotation:** All known nucleotide variants are collected and annotated using tools such as VariantAnnotation. This annotation details the functional consequences of each variant, providing a foundational dataset for subsequent analyses.
2. **Probability Estimation:** We estimate the frequency of each variant using population data from databases like gnomAD. Additionally, we calculate probabilities for random, novel variants that may not yet be catalogued.
3. **Incorporation of Prior Information:** The model integrates biological data (e.g., clinical pathogenicity from ClinVar and structural data from UniProt) as Bayesian priors. This step enhances the model's ability to provide accurate estimates by including comprehensive background information.
4. **Bayesian Inference:** Our approach updates the probability of each variant causing specific diseases by applying Bayesian inference techniques. We use statistical packages such as brms and rstan for this purpose, employing Bayes' theorem to combine prior knowledge with the observed data, resulting in a detailed set of posterior probabilities.

This quantitative expression of disease association likelihood for each variant allows for preemptive strategies in pediatric care tailored to the genetic profiles of individual patients, significantly transforming the landscape of pediatric healthcare. The DeepInfeR (v1.0) initiative not only broadens our understanding of genetic influences on disease but also markedly enhances our capacity to intervene effectively before clinical manifestations occur.

By continuing to refine and expand the DeepInfeR (v1.0) initiative, we anticipate it will become a cornerstone in our predictive capabilities, providing new insights into the genetic basis of diseases and enhancing our ability to offer personalized care.

15 Benefit analysis

15.1 Comparable benchmarks

The following studies are a subset of benchmarks that show the potential of precision medicine, particularly through genomic sequencing and multi-omic technologies, in diagnosing and managing rare and complex conditions. By implementing known methods, we can improve patient outcomes, enable targeted therapies, and potentially reduce healthcare costs through more accurate and faster diagnoses. This integration promises not only to improve patient care but also to provide critical insights into the genetic basis of diseases, ultimately informing both treatment and prevention strategies.

Economic impact of sepsis - [11]

- **Patient data and impact**: Review covers studies reporting costs associated with adult sepsis patients globally, indicating significant economic burden.
- **Methodology**: Systematic review under PRISMA guidelines, with data from PubMed, EMBASE, and Cochrane.
- **Key statistics**: Reports wide cost range from €1,101 to €91,951 per sepsis patient, reflecting international healthcare system variances.

Genomic insights in critical care for infants - [12]

- **Patient data and impact**: Involves 278 critically ill infants; molecular diagnosis achieved in 36.7% of cases, with higher rates (50.8%) in critical trio exome cases.
- **Methodology**: Utilises proband exome, trio exome, and critical trio exome sequencing.
- **Key statistics**: Impacted medical management in 52.0% of diagnosed cases.

National scale multi-omics for rare diseases - [3]

- **Patient data and impact**: Involves 290 critically ill infants and children; diagnostic yield from WGS initially at 47%, increased to 54% with extended analysis.
- **Methodology**: Rapid whole-genome sequencing integrated with transcriptomic data.
- **Key statistics**: Altered critical care management in 77% of diagnosed cases.

Genomic lifespan association in Iceland - [13]

- **Patient data and impact**: Study includes 57,933 participants, identifying 2,306 individuals with actionable genotypes.
- **Methodology**: Whole-genome sequencing focusing on 73 genes from ACMG Secondary Findings recommendations.
- **Key statistics**: Actionable genotypes linked to a decrease in median lifespan by approximately three years for carriers.

Genome analysis in neurodevelopmental disorders - [14]

- **Patient data and impact**: Encompasses 465 families, finding causal variants in 36% of 489 affected individuals.
- **Methodology**: Combines short-read and long-read genome sequencing.
- **Key statistics**: Long-read sequencing crucial for resolving complex variants.

Rapid whole-genome sequencing in UAE - [4]

- **Patient data and impact**: Initial feasibility study with five infants, three of whom were successfully diagnosed.
- **Methodology**: Rapid whole-genome sequencing with a turnaround of 37 hours on average.

- **Key statistics**: Demonstrates the utility of rWGS in diverse populations within a Middle-Eastern context.

Genome sequencing for rare diseases - [2]

- **Patient data and impact**: Involves 822 families with rare monogenic diseases, achieving a diagnostic yield of 29.3%.
- **Methodology**: Focuses on broader genomic coverage including structural and non-coding variants.
- **Key statistics**: Identifies causative variants in 8.2% of cases previously undetected by exome sequencing.

UK and Ireland genomic diagnostics in paediatrics - [1]

- **Patient data and impact**: The DDD study involved over 13,500 families; approximately 41% of probands received a genetic diagnosis.
- **Methodology**: Integrated genomic data analysis with clinical phenotyping.
- **Key statistics**: Highlights cost-saving potential through precise diagnostics and targeted therapy.

15.2 Introduction to source data by Federal Statistical Office

The following data are based on statistics reported by Bundesamt für Gesundheit (BAG), `https://www.bag.admin.ch/` for years 2010-2022. We gathered the yearly statistics for all hospitals and clinics which are gathered for Statistiken zur Krankenversicherung, Qualitätsindikatoren der Schweizer Akutspitäler.

The Bundesamt für Gesundheit (BAG) or Federal Office of Public Health (FOPH), under the mandate of the Federal Department of Home Affairs (FDHA), published quality indicators for hospitals for the first time in 2009 as part of a pilot project. This initiative is based on the revised Federal Health Insurance Act (LAMal), which mandates the publication of data on the quality and cost-effectiveness of service providers.

In collaboration with the Federal Statistical Office (FSO), FOPH chose to adopt the quality indicator concept used by the German HELIOS Clinics. This concept is based on various internationally recognized systems, such as the AHRQ IQIs (Agency for Healthcare Research and Quality - Inpatient Quality Indicators) from the United States and is also utilized by the Quality Medicine Initiative (IQM) in Germany and nationally in Austria. This approach provides Swiss hospitals with a basis for international comparison, particularly benefiting the five Swiss university hospitals by allowing a comparison of their outcomes with those published online from leading German clinics. The indicators are analyzed using existing routine data collected in cooperation with the cantons by the FSO in the Medical Statistics of Hospitals.

With this approach, FOPH enables a systematic, nationwide comparison of outcome quality in acute care hospitals. The Swiss hospital data, which encompasses comprehensive case data in the inpatient sector, is crucial for enabling a thorough comparison based on these routine data. The IQI systems are jointly revised and expanded in Germany and Switzerland. For the current evaluation, the CH-IQI specifications version 5.4 served as the basis for calculations. The indicators provide valuable insights into individual hospital performances and potential areas for improvement, although they do not allow a conclusive judgment on the quality of care provided by hospitals.

15.3 Analysis results for Precision medicine unit

Time series was performed using linear regression (for cost analysis) and Poisson regressions (for mortality to extrapolate the expected outcomes from 2010-2030. The actual cost and number of cases in for the past 12 years is accurately reported by BAG, however or estimate for future savings and benefit depend on the achievable benchmarks reported in the sections of 15. As our number update, we can build a more reliable model of what to expect over the next 10 years. The Precision medicine unit cost analysis overview is shown in figure 8.

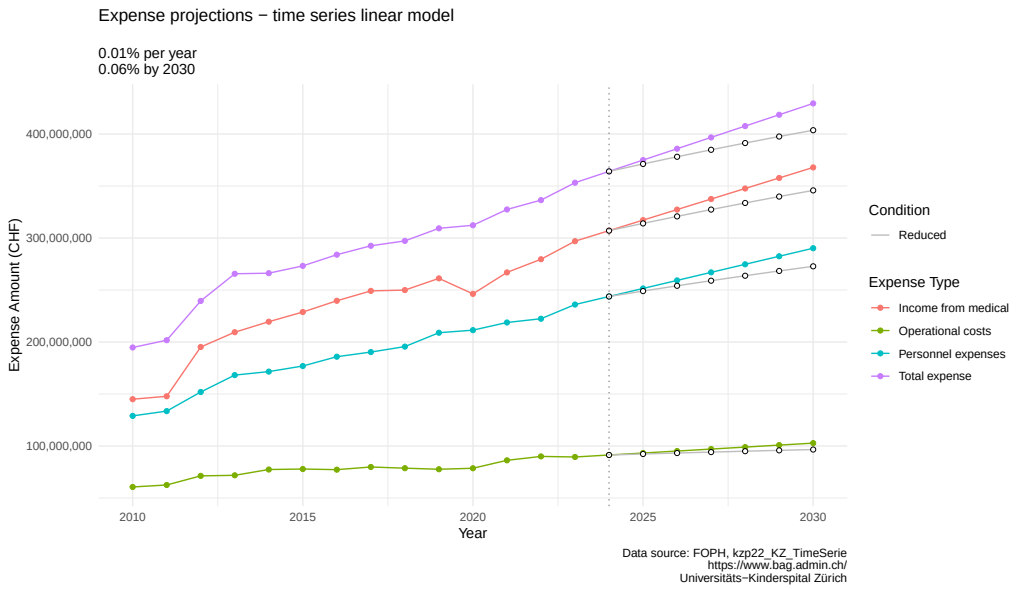

Figure 8: Projections for Kispi with a Precision medicine unit (2010-2030). Federal statistics from Bundesamt für Gesundheit (BAG, Federal office for public health) were modelled and projected to 2030. A modest benefit effect size was applied. In the most similar application to our Precision medicine unit; Lunke et al. [3] showed a 54% diagnostic yield and an altered critical care management in 77% of diagnosed cases. We estimated modest 1% increase in actualised savings per year after successful implementation starting in 2024.

A more accurate demonstration of the Precision medicine unit can be seen with a specific disease example using sepsis. Sepsis was specifically modelled using federal statistics from Bundesamt für Gesundheit (BAG, Federal office for public health) from 2010-2022 as shown in **figures 9, 10, 11**.

First we get a global picture of sepsis in University hospitals in **figure 9**. This illustrates the total number of sepsis in adult and paediatric settings across the country. **Figure 10** shows a forecast model, from 2010-2030, for the number of deaths due to sepsis in Universitäts-Kinderspital Zürich (repeated from executive summary). The predicted number of prevetable deaths is based on comparable benchmarks listed in section 15 which have demonstrated clear cost-saving potential through precise diagnostics and targeted therapy. For rare diseases, approximately 40% of probands received a genetic diagnosis [1; 2] and altered critical care management in 77% of diagnosed cases [3]. A well-managed work-flow can result in rapid whole-genome sequencing with a turnaround of 37 hours on average [4]. Based on such values, the forecast projected into 2030 shows the yearly cases of sepsis. Black and red values show the expected number of deaths with and without precise diagnostics and targeted therapy, respectively.

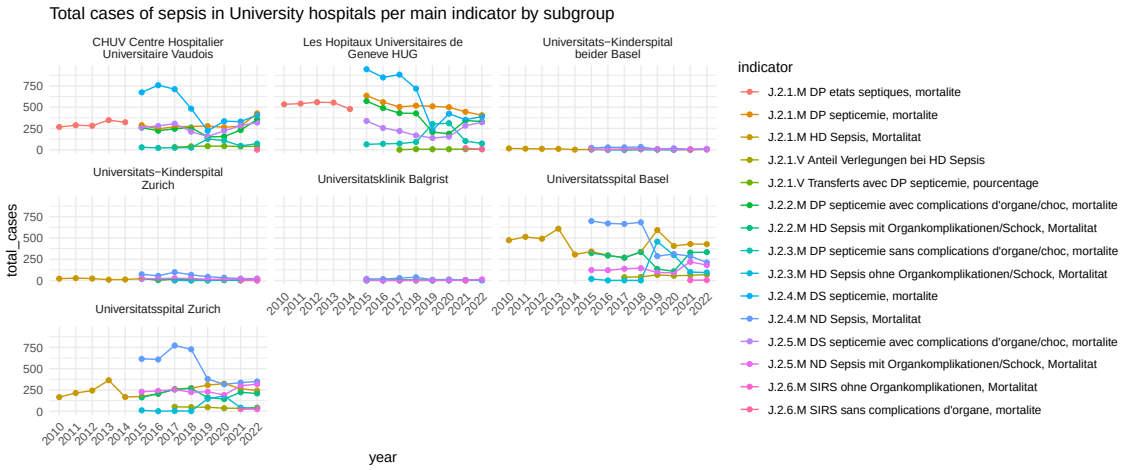

Figure 9: Yearly deaths due to sepsis at University hospitals across Switzerland. This data is based on statistics reported by Bundesamt für Gesundheit (BAG), `https://www.bag.admin.ch/` for years 2010-2022. DP: Diagnostic procedure. HD: Primary diagnosis. ND: Secondary diagnosis.

To put the work of the **Precision medicine unit** in perspective, we look at the total number case statistics for **Universitäts-Kinderspital Zürich**. We see a total number of all cases indicators in 2022 of **10'261**. The subgroup indicator for "J.2 sepsis" shows **74** cases in 2022. **Figure 11** shows the the indicators through A1-Z4, many of which are similarly affected by the advances in precision medicine and are thus potential future prospects.

- J - "Affections complexes, hétérogènes (indicateur pour peer review)" or "Complex, heterogeneous conditions".
  - **J.1** - "Beatmung und extrakorporale Verfahren", or "Artificial respiration and extracorporeal procedures".
  - **J.2** - "Sepsis".
  - **J.3** - "Constellations complexes", or "Multifactorial Conditions".

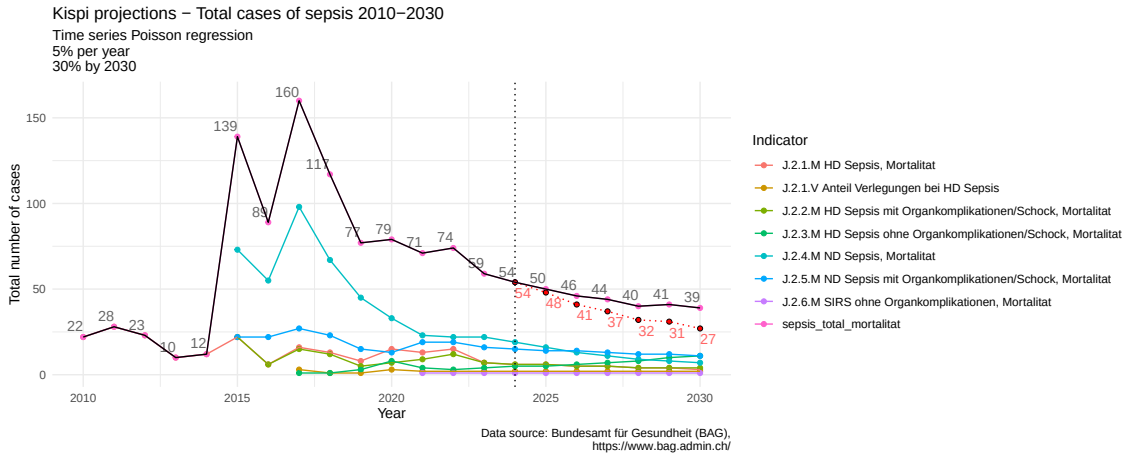

Figure 10: Yearly deaths due to sepsis at Universitäts-Kinderspital Zürich. This data is based on statistics reported by Bundesamt für Gesundheit (BAG), `https://www.bag.admin.ch/` for years 2010-2022. Time series was performed using Poisson regression to extrapulate the expected outcomes from 2010-2030. Predictions for the cost and number of cases were generated in section 15. DP: Diagnostic procedure. HD: Primary diagnosis. ND: Secondary diagnosis.

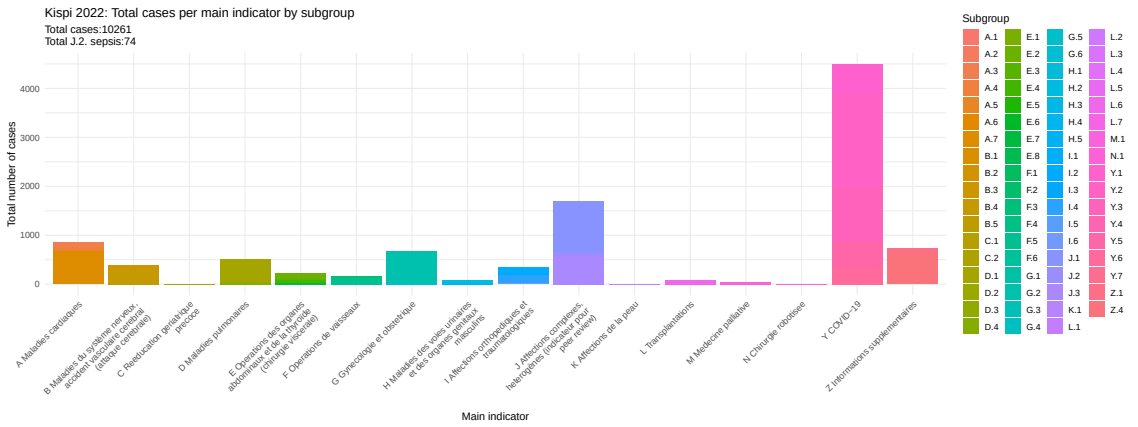

Figure 11: The total count of case descriptions in Universitäts-Kinderspital Zürich for 2022 as categorised according to indicators for the Federal Statistical Office. The 'J.2 Sepsis' subset was used in the other figures shown in this section.

16 Intellectual property

Our approach to intellectual property (IP) diverges from conventional strategies that rely heavily on patents to protect innovations. In the realm of precision medicine, the true value lies not in restricting access to technology through patents but in fostering an environment of open innovation and sharing.

Open-source commitment: We are committed to open-source code and documentation. This commitment supports our mission to advance scientific understanding and innovation in precision medicine. By making our methodologies and technologies publicly available, we enable a collaborative approach where researchers and practitioners can contribute to and benefit from our developments.

Credibility and transparency: We prioritise maintaining high credibility and transparency over the potential financial benefits of patenting our technology. Our stakeholders can value the openness which allows them to verify, adapt, and build upon our work. This transparency is crucial for driving innovation and trust in our processes and outcomes.

Focus on innovation and utility: While we do not dismiss the potential of commercialising discoveries made through our system, our primary focus remains on creating tools and systems that facilitate robust scientific research. We believe that leading through innovation and delivering superior products and services are more crucial for leadership in our field than the exclusive rights granted by patents.

Industry perspective: This philosophy aligns with trends observed in several leading technology companies, where the focus shifts from guarding IP through patents to leading through rapid innovation and open platforms. In these cases, the speed of innovation and the ability to outpace competitors often prove more valuable than traditional patent strategies.

By adopting this IP philosophy, we aim to cultivate a foundation that not only accelerates advancements in precision medicine but also supports a broader scientific community in making impactful discoveries that can lead to better healthcare outcomes worldwide.

17 Conclusion

The Precision medicine unit is purpose-built to meet immediate and achievable healthcare needs by implementing a streamlined and efficient system. This approach circumvents the complexities and delays often associated with incremental changes spread across numerous groups or departments. By centralising efforts with a core team and focusing on integrating advanced omics technologies with current medical practices, the Precision medicine unit can rapidly adapt and apply scientific advancements directly to patient care. This direct implementation pathway ensures that we not only meets current health standards but also sets a robust foundation for future enhancements. Our model of operation emphasises single-source management and open-source advancements. This enables us to deliver reproducible and traceable results that are understandable both internally and externally.

18 Abbreviations

ACAT- aggregated Cauchy association test, AI - artificial intelligence, EMA - European Medicines Agency, GATK - Genome Analysis Toolkit, ICH - International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, IVDR - In Vitro Diagnostic Medical Devices Regulation, MDCG - Medical Device Coordination Group, ML - machine learning algorithms, VSAT - Variant set association testing,

References

[1] Caroline F Wright, Patrick Campbell, Ruth Y Eberhardt, Stuart Aitken, Daniel Perrett, Simon Brent, Petr Danecek, Eugene J Gardner, V Kartik Chundru, Sarah J Lindsay, et al. Genomic diagnosis of rare pediatric disease in the united kingdom and ireland. *New England Journal of Medicine*, 388(17): 1559–1571, 2023.

[2] Monica H Wojcik, Gabrielle Lemire, Eva Berger, Maha S Zaki, Mariel Wissmann, Wathone Win, Susan M White, Ben Weisburd, Dagmar Wieczorek, Leigh B Waddell, et al. Genome sequencing for diagnosing rare diseases. *New England Journal of Medicine*, 390(21):1985–1997, 2024.

[3] Sebastian Lunke, Sophie E Bouffler, Chirag V Patel, Sarah A Sandaradura, Meredith Wilson, Jason Pinner, Matthew F Hunter, Christopher P Barnett, Mathew Wallis, Benjamin Kamien, et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nature medicine*, 29(7):1681–1691, 2023.

[4] Ahmad N Abou Tayoun and Alawi Alsheikh-Ali. A rapid whole-genome sequencing service for infants with rare diseases in the united arab emirates. *Nature Medicine*, 29(12):2979–2980, 2023.

[5] Eelke van der Horst, Deepak Unni, Femke Kopmels, Jan Armida, Vasundra Touré, Wouter Franke, Katrin Crameri, Elisa Cirillo, and Sabine Österle. Bridging clinical and genomic knowledge: An extension of the sphn rdf schema for seamless integration and fairification of omics data. *preprints.org, medicine and pharmacology*, 2023.

[6] Mira N Moufarrej, Diana W Bianchi, Gary M Shaw, David K Stevenson, and Stephen R Quake. Noninvasive prenatal testing using circulating dna and rna: advances, challenges, and possibilities. *Annual Review of Biomedical Data Science*, 6(1):397–418, 2023.

[7] Munir Pirmohamed. Pharmacogenomics: current status and future perspectives. *Nature Reviews Genetics*, 24(6):350–362, 2023.

[8] VLM Yip, DB Hawcutt, and Munir Pirmohamed. Pharmacogenetic markers of drug efficacy and toxicity. *Clinical Pharmacology & Therapeutics*, 98(1):61–70, 2015.

[9] Mary V Relling and William E Evans. Pharmacogenomics in the clinic. *Nature*, 526(7573):343–350, 2015.

[10] Andrea Gaedigk, Scott T Casey, Michelle Whirl-Carrillo, Neil A Miller, and Teri E Klein. Pharmvar: a global resource and repository for pharmacogene variation. *Clinical pharmacology and therapeutics*, 110(3): 542, 2021.

[11] M Van den Berg, FE Van Beuningen, JC Ter Maaten, and HR Bouma. Hospital-related costs of sepsis around the world: A systematic review exploring the economic burden of sepsis. *Journal of Critical Care*, 71:154096, 2022.

[12] Linyan Meng, Mohan Pammi, Anirudh Saronwala, Pilar Magoulas, Andrew Ray Ghazi, Francesco Vetrini, Jing Zhang, Weimin He, Avinash V Dharmadhikari, Chunjing Qu, et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA pediatrics*, 171(12):e173438–e173438, 2017.

[13] Brynjar O Jensson, Gudny A Arnadottir, Hildigunnur Katrinardottir, Run Fridriksdottir, Hannes Helgason, Asmundur Oddsson, Gardar Sveinbjornsson, Hannes P Eggertsson, Gisli H Halldorsson, Bjarni A Atlason, et al. Actionable genotypes and their association with life span in iceland. *New England Journal of Medicine*, 389(19):1741–1752, 2023.

[14] Alba Sanchis-Juan, Karyn Megy, Jonathan Stephens, Camila Armirola Ricaurte, Eleanor Dewhurst, Kayyi Low, Courtney E French, Detelina Grozeva, Kathleen Stirrups, Marie Erwood, et al. Genome sequencing and comprehensive rare-variant analysis of 465 families with neurodevelopmental disorders. *The American Journal of Human Genetics*, 110(8):1343–1355, 2023.

19 Supplemental