Contents

GenomeSwift

Precision medicine in through rapid and automated multi-omic analysis.

1 Summary

GenomeSwift proposes a transformative leap in the diagnosis and treatment of rare genetic diseases through an automated genomic analysis platform, significantly enhancing clinical outcomes for paediatric patients at **UZH** and Universitäts-Kinderspital Zürich. Integrating cutting-edge tools like ProteoMCLustR and ACMGuru, the project promises to streamline fragmented diagnostic processes, ensuring rapid and precise medical insights, starting in metabolic disease. This development aligns perfectly with the Labhart-Schwyzer Scholarship's focus on advancing medical research and technology in metabolic diseases and related fields.

Directed by Dr. Dylan Lawless and supported by collaborations with **UZH**, ETHZ, EPFL, and CHUV, GenomeSwift epitomises interdisciplinary innovation, blending bioinformatics with clinical practice. With a 53-week timeline and focused deliverables, it aims to position **UZH** as a leader in precision medicine, enhancing both the scientific standing of young researchers and the global reputation of Zurich's academic institutions in tackling complex healthcare challenges.

2 Introduction and background

2.1 Multi-omics for diagnosis and discovery to inborn errors of metabolism

We have previously demonstrated the potential for our approach through the SwissPedHealth lighthouse project (`https://www.swisspedhealth.ch`). This stands out as a vanguard in leveraging multi-omics technologies for the diagnosis and discovery of rare metabolic diseases. This initiative has successfully applied genomic, transcriptomic, proteomic, and metabolomic analyses to provide a profound impact on understanding and treating disorders such as Methylmalonic Aciduria (MMA).

Detailed Findings from Phase 1: Methylmalonic Aciduria Study

- **Disease Focus and Genetic Insights:** Focusing on MMA, an inborn error of metabolism, the project identified pathogenic variants in the methylmalonyl-CoA mutase (MMUT) gene in 84% of the studied cases (177 out of 210). This highlights the genetic heterogeneity and complex clinical presentations associated with MMA.
- **Advanced Multi-Omics Methods:** The study utilized a comprehensive suite of multi-omics approaches, integrating whole-genome sequencing (WGS), RNA sequencing (RNA-seq), proteotyping via data-independent acquisition mass spectrometry (DIA–MS), and detailed metabolomic analyses. This integration allowed for a nuanced understanding of the disease at multiple biological levels.
- **Metabolic Pathway Disruptions:** Significant disruptions were discovered in the tricarboxylic acid (TCA) cycle and its anaplerosis, primarily involving glutamine. These disruptions were extensively characterized through multi-organ metabolomics and stable-isotope tracing in a hemizygous Mmut mouse model, providing a clear pathophysiological pathway that contributes to the disease state.
- **Protein Interactions and Therapeutic Insights:** The study underscored crucial interactions between MMUT and key enzymes such as glutamate dehydrogenase and oxoglutarate dehydrogenase. Treatment with dimethyl-oxoglutarate was shown to restore TCA cycle functionality, offering a novel therapeutic avenue that could significantly impact clinical outcomes for patients with MMA.

Ongoing and Future Research Phases

- **Phase 2 and Phase 3:** Building on the findings from Phase 1, ongoing phases aim to extend these multi-omics methodologies to broader cohorts with extreme phenotypes. The project seeks to refine diagnostic workflows and explore the real-time clinical application of these findings in prospective studies involving children with severe metabolic dysfunctions.

2.2 GenomeSwift

The integration of genomic technology into healthcare, especially for rare diseases, represents a crucial frontier in medical science [1]. Despite advancements, the fragmented nature of current genome analysis methodologies in Swiss healthcare notably delays diagnoses and treatment. The accumulation of extensive genomic datasets necessitates a shift towards more integrated and automated processes [2].

Our proposed GenomeSwift bioinformatic suite aims to address these **unmet needs**. It is designed to automate and amalgamate various genome sequence analysis processes, thus **feasibly** providing swift, precise, and expansive insights into genetic diseases. This suite adheres to FAIR data principles, ensuring advancement in research and conformity with best data management practices in Universitäts-Kinderspital Zürich.

The persistent bottleneck in translating genomic data into practical clinical insights underscores the project's necessity [3]. Even with tools like the Genome Analysis Toolkit setting standards for variant analysis, a comprehensive system integrating these tools is essential for a robust analysis process [4]. Our **previous developments** underpin GenomeSwift (**Figure 1**) [5], [6], [7]. We have additionally developed several tools that we are excited to integrate into the pipeline, including:

- ProteoMCLustR, a protocol for protein pathway clustering [8], [7].
- SkatRbrain, a statistical analysis pipeline effective in analysing rare variants [9], [10], [11].
- Archipelago, for unified genomic statistical analysis.
- ACMGuru, incorporating American College of Medical Genetics guidelines to translate genomic data into clinical recommendations [12].
- DeepInferR, for Bayesian analysis of genetic variant probabilities and their disease impacts.

By merging these tools into a unified pipeline, GenomeSwift aims to elevate the efficiency and output of genomic data analysis in Universitäts-Kinderspital Zürich, enabling real-time, high-throughput genomic data analysis for disease discovery.

2.3 Precision medicine unit operational design

The Precision medicine unit operates as a self-organised, start-up group composed of researchers and clinicians from Universitäts-Kinderspital Zürich, UZH, and ETHZ. We apply advanced multi-omics technologies for quick diagnosis and tailored treatment strategies in pediatric care. The core philosophy centres on robust project management systems and cutting-edge bioinformatics pipelines to ensure swift clinical implementation of research findings. Our operational structure is based on a unified system framework that employs a single-source management philosophy to enhance data integration and operational efficiency. We uphold high standards by building for rigorous audit trails and version control systems. Funding will be used to develop bioinformatics tools, expand diagnostic platforms with new omics technologies.

3 Objective and hypotheses

The primary objective of this study is to develop and implement GenomeSwift, an automated, comprehensive software pipeline designed for rapid, precise genomic data analysis, with a particular focus on the diagnosis and treatment of genetic diseases in Universitäts-Kinderspital Zürich.

3.1 Hypotheses:

(i) Automation and integration efficiency: Automating the genome analysis process and integrating various existing tools into a single pipeline, GenomeSwift is expected to significantly expedite genomic data processing, thus enhancing the efficiency and speed of diagnosing and planning treatments for diseases. **(ii) Enhanced diagnostic accuracy**: By incorporating advanced tools for variant detection, statistical analysis, and clinical interpretation, GenomeSwift aims to improve diagnostic accuracy for rare diseases, leading to more precise and personalised treatment approaches. **(iii) Impact on disease research**: The use of GenomeSwift in rare disease contexts is anticipated to not only improve clinical outcomes but also to enrich genomic research, offering a workflow that provides traceable genetic evidence to better classify variants.

3.2 Objectives:

(i) Tool integration Integrate existing tools such as ProteoMCLustR, SkatRbrain, and ACMGuru into a unified and automated pipeline to streamline the process from data input to clinical interpretation. **(ii) Validation and refinement**: Validate GenomeSwift using both simulated [13] and real-world datasets to confirm its reliability and accuracy across various clinical scenarios, refining the pipeline based on performance metrics and feedback [14], [15]. **(iii) User accessibility**: Ensure GenomeSwift is user-friendly for analysts while making evidence-based results accessible to healthcare professionals and researchers, facilitating broader adoption within Universitäts-Kinderspital Zürich and potentially other institutions. **(iv) Knowledge dissemination**: Disseminate the findings and capabilities of GenomeSwift through publications and collaborations, aiming to enhance the use of genomic data for healthcare improvement and scientific discovery.

Figure . From DNA to diagnosis: Variant effect evidence is assessed based on standardised guidelines. Data is produced in an analysis-friendly formats for statistical or AI/ML reuse. GenomeSwift produces clinical genetics reports and database results using SPHN RDF schema concepts.

4 Materials and methods

The GenomeSwift pipeline is designed to enhance genomic analysis through automation and integration of various analytical tools. Our detailed research plan outlines the methodology and statistical approaches we will employ to ensure the effectiveness and efficiency of GenomeSwift in processing and analysing genomic data.

(i) Integration of existing tools: The pipeline will integrate several tools that we have developed, including: ProteoMCLustR for protein pathway clustering [7], [8]; SkatRbrain for statistical analysis of genetic data [9], [10], [11]; Archipelago for a unified representation for genomic statistical analysis; ACMGuru for clinical genetic interpretation [12]; AutoConstructR for protein structure plotting, facilitating a comprehensive interpretation of genetic variations [13]; and other modular tools.

(ii) Data processing and analysis workflow: Data input: GenomeSwift will accept raw genomic data, applying preprocessing steps to ensure data quality and compatibility. **Variant detection**: Utilising the best practices from tools like GATK, the pipeline will perform variant calling, ensuring high-confidence

identification of genetic variations [4]. **Statistical analysis**:Employing SkatRbrain, the pipeline will conduct robust statistical analyses to associate genomic variations with disease phenotypes, including rare variant analysis [9], [10], [11]. **Clinical interpretation**: ACMGuru will be used to interpret the clinical significance of detected variants, aligning with the American College of Medical Genetics guidelines [12].

(iii) Simulation and validation: The pipeline's efficacy will be validated using simulated datasets encompassing various disease scenarios (rare variant, common variant, polygenic risk) to ensure its robustness across different genetic contexts [3], [14], [15]. Validation will also include real-world data from Swiss hospitals to confirm the pipeline's practical applicability and accuracy.

(iv) Statistical methodologies: GenomeSwift will incorporate advanced statistical methods to analyse the association between genetic variants and diseases, ensuring the analyses are powered adequately to detect significant associations even in the context of rare diseases. The pipeline will employ a range of statistical tests suitable for different data types and study designs, ensuring the flexibility and comprehensiveness of the analysis. Specifically, optimised sequence kernel association tests (SKAT-O) will form the basis of statistical validation tests [10]. Successful outcomes will therefore demonstrate the ability to substitute compatible drop-in methods; burden tests such as CMC [16] and WSS [17], variance component tests such as C-alpha [18] and SKAT [9], combined burden and variance component tests such as SKAT-O [10], other combination tests such as ACAT-RVAT [19], regression and generalised mixed models such as REGENIE [20] and SAGE-GENE+ [21], and others.

(v) Automation and user interface: The pipeline will feature containerisation to support development and use. Automation will be a key focus, with the pipeline designed to require minimal user intervention, streamlining the analysis process from start to finish. User output will include graphical interfaces and technical reporting documents.

(vi) Output and reporting: GenomeSwift will generate comprehensive reports, detailing the analysis results, including variant identification, statistical associations, and clinical interpretations. The pipeline will ensure that outputs are presented in an easily interpretable format, facilitating clinical decision-making and further research. The key technical data will also be generated including formats for reporting with SPHN RDF schema concepts.

5 Expected outcomes and significance

GenomeSwift represents a pivotal development at the intersection of advanced genomics and clinical application, poised to influence diagnosis and treatment protocols. Its significance spans several critical aspects for Universitäts-Kinderspital Zürich:

- **Diagnostic and Therapeutic Advancements:** The project not only enhances the diagnostic precision for rare metabolic diseases like MMA but also facilitates the development of targeted therapeutic strategies based on deep molecular insights.
- **Contribution to Metabolic Disease Research:** By providing a comprehensive molecular map of MMA and demonstrating effective intervention strategies, the Lighthouse Project contributes significantly to the field of metabolic disease research, setting new standards for clinical practice.
- **Healthcare Impact:** GenomeSwift will enhance the diagnostic process for rare diseases by delivering rapid and precise genomic analysis. This capability enables more timely and **accurate treatment decisions**, crucial for improving patient outcomes. By providing detailed genetic insights, GenomeSwift supports the shift towards **personalised medicine**, where treatments are specifically tailored to individual genetic profiles.

- **Strategic alignment:** GenomeSwift aligns with the strategic objectives of the Children's Hospital Zurich, notably enhancing the institution's **capacity** for genetic diagnostics and personalised healthcare. The project also contributes to the efficiency and effectiveness of the **broader Swiss healthcare system**, underscoring innovation and leadership in medical genomics.
- **Scientific advancement:** This project marks a considerable advance by amalgamating multiple analytical tools into a unified pipeline, thereby establishing a new benchmark in genomic analysis. GenomeSwift will streamline genomic data processing, **accelerating** research into rare diseases and potentially revealing novel therapeutic targets.
- **Collaboration and Education:** GenomeSwift will enhance **collaboration** across researchers, clinicians, and institutions by offering a shared genomic analysis format, fostering an integrated healthcare and research approach. Additionally, it will serve as an **educational resource** to improve genomic literacy and train future experts, clarifying the application of clinical genetics guidelines in analysis for healthcare professionals.
- **Sustainability and accessibility**: As an open-source initiative, GenomeSwift is accessible to a diverse user base, promoting ongoing support. The project's focus on understanding and treating diseases has the potential to make a **lasting impact** on healthcare and research, leading to enhanced patient outcomes and broader scientific insights.

5.1 Added value for the applicant's career and home institution

This project represents a key advancement in my early scientific career, positioning me as the lead for integrating advanced research with translational clinical applications. As the responsible developer of GenomeSwift - the inaugural product from the Precision medicine unit- I am building a novel approach to translational research in medicine. This role will enhance my leadership skills to meet the expectations of the Precision medicine unit. Simultaneously, managing GenomeSwift's deployment in real-time, critical care settings will help me lay a foundation for innovative disease treatment methodologies.

This initiative will demonstrate **UZH**'s proficiency for creating and managing complex datasets, thereby becoming an essential asset for a broad range of rare disease, starting in metabolism, and planed to expand into immunology, endocrinology, and diabetology.

By spearheading the adoption of innovative omic analysis techniques, this project promises to markedly elevate both my professional growth and **UZH**'s influence in the realm of precision medicine. Further details are provided in our whitepaper: < link to github.com PDF download>.

5.2 Value analysis

5.2.1 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.

The economic impact of critical care [22] reveals substantial costs per patient (systematic review: €1,101 to €91,951). In genomic insights for critically ill infants [23], molecular diagnoses affected medical management in over half the cases, utilizing critical trio exome sequencing. National scale multi-omics

approaches for rare diseases [24] improved diagnostic yields to 54% with integrated rapid whole-genome and transcriptomic data, significantly altering critical care management in 77% of cases. A study on genomic lifespan associations in Iceland [25] identified actionable genotypes using whole-genome sequencing, which correlated with a decrease in median lifespan by three years. In the realm of neurodevelopmental disorders [26], combined short-read and long-read sequencing identified causal variants in 36% of the cases, highlighting the importance of long-read technology in resolving complex variants. Rapid whole-genome sequencing in the UAE [27] demonstrated its utility with a quick turnaround, particularly noting its effectiveness in a Middle-Eastern population. Extensive genome sequencing for rare diseases [28] achieved a 29.3% diagnostic yield, significantly identifying causative variants previously undetected by exome sequencing. Lastly, the UK and Ireland genomic diagnostics in paediatrics [29] through the DDD study showed a 41% diagnostic rate, emphasizing cost savings and targeted therapeutic potentials.

5.3 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. 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 1.

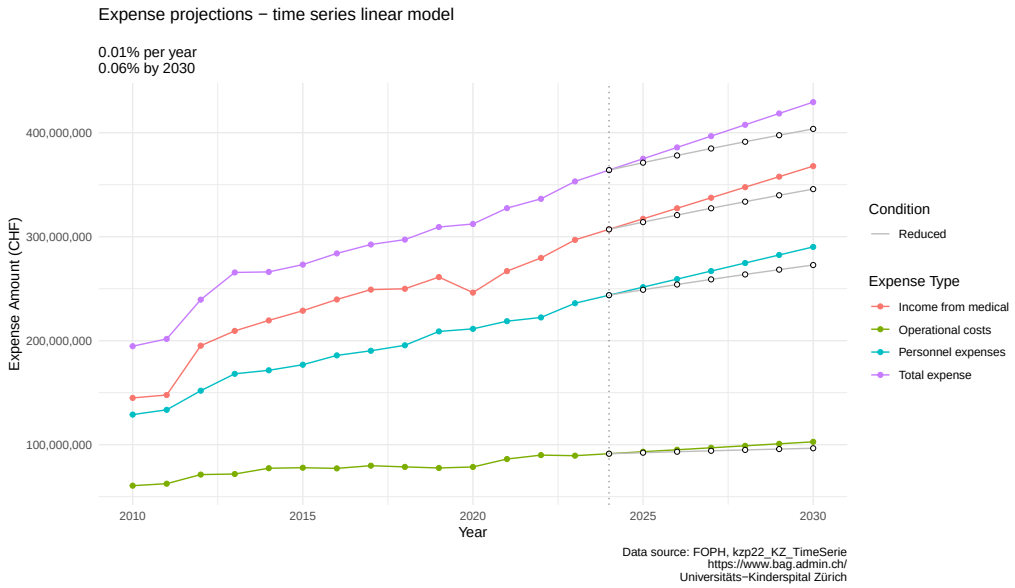

Figure 1: Projections for Kispi with a Precision medicine unit (2010-2030). Federal statistics from Bundesamt für Gesundheit (BAG, Federal office for public health) were modelled and projected to 2030. A modest benefit effect size was applied. In the most similar application to our Precision medicine unit; Lunke et al. [24] 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 **figure** 2. The forecast model, from 2010-2030, shows the number of deaths due to sepsis in Universitäts-Kinderspital Zürich (**Figure** 2). The predicted number of preventable deaths is based on comparable benchmarks listed in section 5.2.1 which

have demonstrated clear cost-saving potential through precise diagnostics and targeted therapy. For rare diseases, approximately 40% of probands received a genetic diagnosis [28; 29] and altered critical care management in 77% of diagnosed cases [24]. A well-managed work-flow can result in rapid whole-genome sequencing with a turnaround of 37 hours on average [27]. 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.

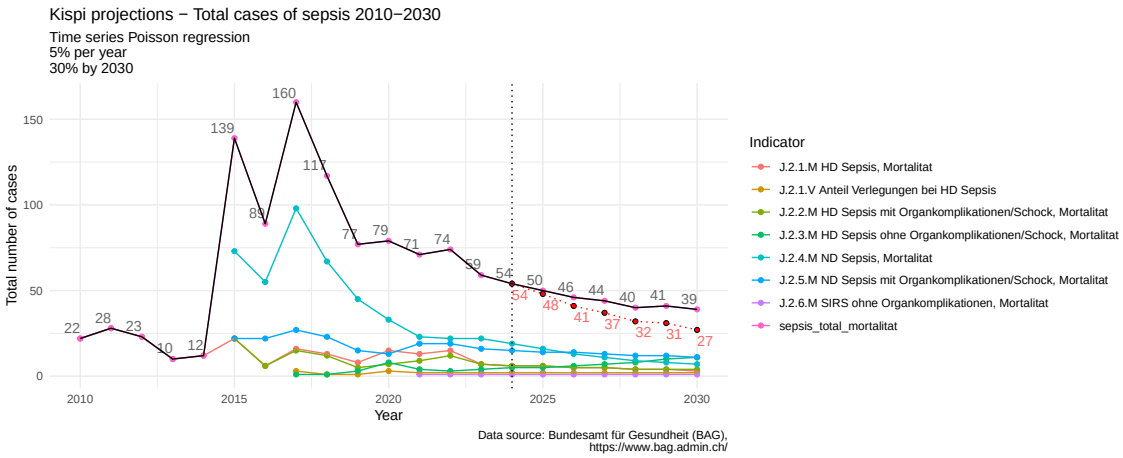

Figure 2: Yearly deaths due to sepsis at Universitäts-Kinderspital Zürich. This data is based on statistics reported by Bundesamt für Gesundheit (BAG), `https://www.bag.admin.ch/` for years 2010-2022. Time series was performed using Poisson regression to extrapulate the expected outcomes from 2010-2030. Predictions for the cost and number of cases were generated in section 5.2.1. 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.

6 Collaborations and network

6.1 Involved persons

The GenomeSwift project is supported by a multidisciplinary team of experts, each bringing unique expertise and experience to ensure the project's success. **Prof. Luregn Schlapbach** will act as **mentor**: project management, expected milestones, and outcomes will be tracked with progress reports. **Prof. Jacques Fellay** will provide **advice** on the tool requirements, essential guidance on data privacy, consent, and ethical considerations, ensuring that GenomeSwift adheres to the highest standards of research ethics and data protection. **Dr. Dylan Lawless** will act as **project leader**: established in genomics and bioinformatics, with extensive experience in developing computational tools for genomic analysis. He has expertise in the understanding genetic variations and their implications in diseases, especially in the context of rare disease in children. **Dr. Vito Zanotelli** will act as bioinformatics specialist: he has a profound understanding of integrating and analysing large-scale genomic datasets. His expertise is crucial in refining the data processing algorithms and ensuring that GenomeSwift can handle complex and voluminous datasets efficiently. **Ali Saadat** has a background in genetic analysis method development, bringing valuable insights into the clinical implications of genomic findings. His focus is on translating genomic data into actionable clinical knowledge, which is instrumental in designing the interpretative aspects of GenomeSwift. **Dr**

Zhi Ming Xu will advise as genomic analysis statistics and software development. A **PhD student** will be employed to work on development of novel research features.

Additional collaborating bioinformatics specialists from UZH, ETHZ, EPFL, and CHUV will integrate and optimise the computational tools within GenomeSwift. They will test the flow of data in the format matching our data providers (1) (e.g., SwissMultiOmic Center) to (2) HPC clusters (e.g., BioMedIT) into (3) GenomeSwift. They will test the clinical applicability and relevance of the pipeline. Open-source development will be tested by our collaborators in swisspedhealth.ch, and EPFL. Software will be tested on multiple nodes including ETHZ SIS Leonhard Med and University of Basel sciCORE Med.

6.2 National / international collaborations

GenomeSwift is enhanced by an extensive network of collaborations both nationally and internationally, significantly impacting healthcare and research communities. These partnerships facilitate knowledge sharing and resource exchange, crucial for the project's success. **SwissMultiOmic Center:** Serving as our primary data provider, this key national partner offers access to state-of-the-art technologies and datasets, enabling GenomeSwift to utilise high-quality genomic data and analytical resources. Regular communication between our groups aids in refining and advancing the pipeline. **Centre Hospitalier Universitaire Vaudois (CHUV):** Collaboration with CHUV researchers to review GenomeSwift's adaptability across different healthcare settings, ensuring its effectiveness and versatility. CHUV's commitment to medical genetics research provides a solid basis for collaborative enhancements of the pipeline. **Ecole Polytechnique Fédérale de Lausanne (EPFL) and ETH Zurich:** These partnerships grant GenomeSwift access to extensive expertise in bioinformatics, computational biology, and genomics, fostering a multidisciplinary integration of varied user needs and methodologies. **Global Alliance for Genomics and Health (GA4GH):** GenomeSwift aligns with GA4GH standards to enhance data interoperability and security worldwide (`https://www.ga4gh.org`). This partnership ensures GenomeSwift's integration into global genomic databases, adhering to international best practices in data privacy and ethics, thereby contributing to the global 'internet of genomics'. **Open-source software community:** As a collaborative open-source project, GenomeSwift benefits from the collective insights of a diverse group of developers and researchers, promoting continuous innovation and ensuring the platform's ongoing availability and improvement.

7 Timeline and milestones

Our project timetable was designed and assessed using the critical path method (CPM) and program evaluation and review technique (PERT). In **Figure 2** (A) we present the network of project activities, highlighting critical paths crucial for project completion. **Figure 2** (B) illustrates the Gantt chart, detailing activity timelines and critical paths, which we tested for effective project scheduling. Lastly, **Figure 2** (C) shows the probabilistic distribution of project completion time, where we tested the risk and probability of meeting deadlines. Our analysis using CPM and PERT, provides the optimal plan to assist in timely delivery of outcomes. With a calculated probability of completion within 53 weeks at 0.97, the project is highly likely to finish on time according to the PERT analysis [30], [31], [32], [33], [34]

Figure . Timetable with PERT network and Gantt chart. (A) Network diagram depicting project activities and critical paths. Critical paths with time dependencies highlighted by color. (B) Gantt chart illustrating project timelines and critical paths. (C) Probability distribution of project completion time for risk assessment and deadline management. CR, time-critical; NC, non-critical.

8 Financial plan and budget

8.1 Available resources

The GenomeSwift project will use existing resources and infrastructure available through our collaborations and institutional support, ensuring a cost-effective approach to its development and implementation. **Institutional support**: Access to computational resources and infrastructure provided by our collaborating institutions, including the SwissMultiOmic Center, ETHZ SIS Leonhard Med, and University of Basel sciCORE Med. **Existing grants**: We will also performed development as part of current funding from related projects within our group which can be found at their respective research project pages: swisspedhealth.ch, and EPFL, CHUV.

8.2 Requested resources

8.2.1 Personal costs (including social security contributions)

A senior staff scientist and a PhD student will each be employed 50% FTE to work on development, including social security contributions. Additional funding is not required for the following: the main applicant who will oversee the project, collaborating bioinformatics specialists from UZH, ETHZ, EPFL, and CHUV who will integrate and optimise the computational tools within GenomeSwift under their current roles within their respective institutions; our collaborating clinical geneticists who will review the clinical applicability and relevance of the pipeline. Total personal costs are 90’448 CHF.

8.2.2 Material costs

Costs associated with high-performance computing (HPC) resources for data processing and analysis are listed in Table 1, including discounts covered by institutional support. Total costs are 7’628.

8.3 Summary budget table

1 year project costs	
Senior staff (107'729) 50% FTE	53865
Social security contributions (estimated 16%)	8618
Postdocs (i.e. 91'280)	0
Salary for doctoral student (i.e. 48'216) 50% FTE	24108
Social security contributions (estimated 16%)	3857
Other	
Total Direct Costs for Personnel	90448
Material costs	
HPC GPU time	1520
HPC compute time	2008
HPC storage 30 TB	2100
HPC configuration and support	1000
HPC access services	1000
Total material costs	7628
Total	98076

9 References

References

[1] Jean-Laurent Casanova and Laurent Abel. The human genetic determinism of life-threatening infectious diseases: genetic heterogeneity and physiological homogeneity? *Hum. Genet.*, 139:681–694, 2020.

[2] Gundula Povysil, Slavé Petrovski, Jason Hostyk, Vandana Aggarwal, A. S. Allen, and David B. Goldstein. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat. Rev. Genet.*, 20(12), 2019. doi: 10.1038/s41576-019-0177-4.

[3] B. S. Pedersen et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease. *NPJ Genomic Med.*, 6(1):1–8, 2021.

[4] G. A. Van der Auwera et al. From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.*, 43(1):11.10.1–11.10.33, 2013. doi: 10.1002/0471250953.bi1110s43.

[5] D. Lawless et al. Prevalence of cftr variants in primary immunodeficiency patients with bronchiectasis is an important modifying cofactor. *J. Allergy Clin. Immunol.*, 152(1):257–265, 2023. doi: 10.1016/j.jaci.2023.01.035.

[6] D. Lawless et al. Viral genetic determinants of prolonged respiratory syncytial virus infection among infants in a healthy term birth cohort. *J. Infect. Dis.*, 227(10):1194–1202, 2023. doi: 10.1093/infdis/jiac442.

[7] Z. M. Xu et al. Genome-to-genome analysis reveals associations between human and mycobacterial genetic variation in tuberculosis patients from tanzania. *Published online*, 2023. doi: 10.1101/2023.05.11.23289848.

[8] D. Szklarczyk et al. The string database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.*, 45(D1):D362–D368, 2016. doi: 10.1093/nar/gkw937.

[9] M. C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, 89(1):82–93, 2011. doi: 10.1016/j.ajhg.2011.05.029.

[10] S. Lee et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.*, 91(2):224–237, 2012. doi: 10.1016/j.ajhg.2012.06.007.

[11] S. Lee, M. C. Wu, and X. Lin. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, 13(4):762–775, 2012. doi: 10.1093/biostatistics/kxs014.

[12] S. Richards et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genet Med*, 17(5):405–424, 2015. doi: 10.1038/gim.2015.30.

[13] The UniProt Consortium. Uniprot: the universal protein knowledgebase. *Nucleic Acids Res.*, 45(D1):D158–D169, 2016. doi: 10.1093/nar/gkw1099.

[14] X. Li et al. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat. Genet.*, 52(9):969–983, 2020. doi: 10.1038/s41588-020-0676-4.

[15] Z. Li et al. A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nat. Methods*, 19(12):1599–1611, 2022. doi: 10.1038/s41592-022-01640-x.

[16] B. Li and S. M. Leal. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.*, 83(3):311–321, 2008.

[17] B. E. Madsen and S. R. Browning. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*, 5(2):e1000384, 2009. doi: 10.1371/journal.pgen.1000384.

[18] B. M. Neale et al. Testing for an unusual distribution of rare variants. *PLoS Genet.*, 7(3):e1001322, 2011.

[19] Y. Liu, S. Chen, Z. Li, A. C. Morrison, E. Boerwinkle, and X. Lin. Acat: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.*, 104(3): 410–421, 2019.

[20] J. Mbatchou et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.*, 53(7):1097–1103, 2021. doi: 10.1038/s41588-021-00870-7.

[21] W. Zhou et al. Saige-gene+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.*, 54(10):1466–1469, 2022. doi: 10.1038/s41588-022-01178-w.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] D. Miszczyńska and M. Miszczyński. *Wybrane metody badań operacyjnych*. Published independently, 2002.

[31] P. R. Murthy. *Operations research (linear programming)*. bohem press, 2005.

[32] Y. Cohen and A. Sadeh. A new approach for constructing and generating aoa networks. *J. Eng. Comput. Archit.*, 1(1):1–13, 2007.

[33] H. A. Taha. *Operations research an introduction*. Published independently, 2007.

[34] I. Konarzewska, M. Jewczak, and A. Kucharski. *Optymalizacja w logistyce, tom 1. Modelowanie logistycznych procesów decyzyjnych*. Wydawnictwo Uniwersytetu Łódzkiego, 2020.