MoBa Explorer: Enabling the Navigation of Data from the Norwegian Mother, Father and Child Cohort Study (MoBa)

H. Balaka[1,2,3], R. Valen [1], L. A. Garrison[2] and M. Vaudel[1,3,4]

[1]Norwegian Institute of Public Health, Department of Genetics and Bioinformatics, Health Data and Digitalization, Norway
[2]University of Bergen, Department of Informatics, Visualization Group, Norway
[3]University of Bergen, Department of Clinical Science, Mohn Center for Diabetes Precision Medicine, Norway
[4]University of Bergen, Department of Informatics, Computational Biology Unit, Norway

Abstract
Studies in public health have generated large amounts of data helping researchers to better understand human diseases and improve patient care. The Norwegian Mother, Father and Child Cohort Study (MoBa) has collected information about pregnancy and childhood to better understand this crucial time of life. However, the volume of the data and its sensitive nature make its dissemination and examination challenging. We present a work-in-progress design study and accompanying web application, the MoBa Explorer, which presents aggregated MoBa study data genotypes and phenotypes. Our research explores how to serve two distinct purposes in one application: (1) allow researchers to interactively explore MoBa data to identify variables of interest for further study and (2) provide MoBa study details to an interested general public.

CCS Concepts
• *Human-centered computing → Visualization;*

1. Introduction

In the realm of public health, a substantial corpus of data has been collected offering profound insights into human disease trajectories and comorbities. One such study is the Norwegian Mother, Father and Child Cohort Study (MoBa) [MBV*16], which captures pivotal information concerning pregnancy and early childhood. Effective presentation of this data can help researchers hypothesize and verify their theories. Additionally, accessible presentation of this information can aid in its dissemination to the general public with varying degrees of health literacy. However, MoBa study data are complex, heterogeneous, and contain sensitive patient information. As such, MoBa data are largely inaccessible to researchers and the general public to explore and use. Researchers need to access MoBa data to identify variables of interest for study requisition, while lay audiences, many of whom may have been study participants, would like to access the data to, e.g., see how their data are being used or to understand how likely a person with their, or their child's profile, may be to inherit certain diseases or conditions. Visualization is a proven way to facilitate user understanding and task performance with complex datasets [Mun15], through visual analysis pipelines or through targeted presentation of information.

Several visualization approaches have been developed to support exploratory analysis of genomic [NHG19] and phenotype data [PXH*18], with specialized toolkits like Gosling for visualizing these data [LWLG21]. These approaches are developed mainly for experts. While projects such as FinnGen [fin] and UK Biobank [ukb] provide public-facing visual web interfaces to explore genomic and phenotype cohort data similar to those from MoBa, these interfaces require advanced domain expertise, and are in practice unapproachable to a broader public. Accessibility is now a legal requirement for public services in Norway, and the development of approaches for accessible visualizations of complex data is an active area of research in visualization [EBM22]. The key and unique challenge in our project is to identify the needs and use cases of two disparate user groups, and to find an intersection of visualization methods for genomic and phenotypic data that can serve the most important needs of researchers and of an interested general public. This poster presents our research-in-progress and next steps for a design study and prototype web application, the MoBa Explorer, to address this challenge.

2. Data, Tasks & Requirements

The Norwegian Mother, Father, and Child Cohort Study (MoBa) is an ongoing longitudinal study by the Norwegian Institute of Public Health (NIPH) of the causes of disease among mothers and children, with input from fathers [MBV*16]. Beginning in 1999, the study has collected information from over 100,000 pregnancies. The data include twelve collected time points over several years and are mixed-type (qualitative and quantitative). Data collected include biological material (e.g., blood, DNA, placental) as well as

qualitative questionnaires to uncover dietary, environmental, and lifestyle habits. MoBa data are stored across disparate questionnaires and databases that are difficult to search.

Tasks. Researchers using MoBa data wish to identify study variables of interest for requisition and further investigation, e.g., variables influencing early-childhood growth. For each variable, they must identify the number of individuals expressing that variable, and the percentage of missingness. They wish to link gene expression variables to resulting phenotype(s) variables. A researcher is also interested in exploring how variable expression changes over time, or differs according to social or demographic factors. Finally, they wish to compare expression differences between related mothers, fathers, and children. A lay user is interested in the former points, but from a more personal perspective: they wish to understand how their personal health information, or that of their parents (if they were a child in the study) or how their children (if their child was in the study) compares to others in the study.

Requirements. The main artifact of this research is a web application that will enable researchers and the general public to access MoBa study data in collaboration with the Norwegian Institute for Public Health. The application should unify the disparate data sources processed through projects like MoBaPsychGen [CFS*22], enabling users to search for and access variables from the MoBa study. Data should be possible to subset according to key demographic features of interest, and display changing values of variables over time. Data cannot be individually identifiable, and must be presented in aggregate. The visualization should include an option to show missingness of information for a given variable. To adhere to Norwegian law, the application must meet accessibility standards.

3. MoBa Explorer & Next Steps

We have developed a proof-of-concept MoBa Explorer addressing some requirements identified in the previous section. The interface provides a list of time points and corresponding variables encoded as an accordion dropdown so that a user can select a time point and then proceed to a variable of interest. For each variable depicted by descriptive statistics, the interface shows flat tables with statistical attributes giving exact numbers and boxplot charts, so that a user can compare subsets of the data (Figure 1). As we are limited by descriptive statistics, box-and-whisker diagrams give a concise and constructive way to visualize this data [Mun15]. For variables with categorical attributes, we are exploring use of stacked bar charts to display categories in a part-whole relationship. While displaying a variable, our interface lets users interact with it by selecting subsets and focusing on them [Hau06].

This project is in its early stages with much to be done. We have thus far focused mainly on expert interviews and tasks, but must talk to lay users to understand their needs. The current visual encodings are a first step, but we are exploring alternatives that may show these data in an equally or more expressive and engaging manner. Critical to this process is to explore intuitive ways to make the data visually accessible, through aria labels and other means to "chunk" the data for a screen reader. Finally, we plan to conduct pilot studies and evaluations of the tool to asess its success across expert and lay audience groups.

Figure 1: *The MoBa Explorer Prototype consists of four panels. In the top panel, users may choose About to learn about the MoBa study, or to browse and select MoBa Genotypes or Phenotypes, or search through the search panel. The left panel allows selection of time points. The central panel shows basic statistics for each of the categories within a selected variable, showed in tabular form for accessibility purposes and in a box plot below. The right panel enables filtering of these categories.*

References

[CFS*22] CORFIELD E., FREI O., SHADRIN A., RAHMAN Z., LIN A., ATHANASIU L., AKDENIZ B., HANNIGAN L., WOOTTON R., AUSTERBERRY C., HUGHES A., TESLI M., WESTLYE L., STEFANSSON H., STEFANSSON K., NJØLSTAD P., MAGNUS P., DAVIES N., APPADURAI V., HAVDAHL A.: The norwegian mother, father, and child cohort study (moba) genotyping data resource: Mobapsychgen pipeline v.1, 06 2022. doi:10.1101/2022.06.23.496289. 2

[EBM22] ELAVSKY F., BENNETT C., MORITZ D.: How accessible is my visualization? evaluating visualization accessibility with chartability. In *Computer Graphics Forum* (2022), vol. 41, Wiley Online Library, pp. 57–70. doi:10.1111/cgf.14522. 1

[fin] FinnGen results — r7.finngen.fi. https://r7.finngen.fi/. [Accessed 07-09-2023]. 1

[Hau06] HAUSER H.: *Generalizing Focus+Context Visualization*. 01 2006, pp. 305–327. doi:10.1007/3-540-30790-7_18. 2

[LWLG21] L'YI S., WANG Q., LEKSCHAS F., GEHLENBORG N.: Gosling: A grammar-based toolkit for scalable and interactive genomics data visualization. *IEEE Transactions on Visualization and Computer Graphics PP* (10 2021), 1–1. doi:10.1109/TVCG.2021.3114876. 1

[MBV*16] MAGNUS P., BIRKE C., VEJRUP K., HAUGAN A., AL-SAKER E., DALTVEIT A. K., HANDAL M., HAUGEN M., HØISETH G., KNUDSEN G. P., ET AL.: Cohort profile update: the Norwegian mother and child cohort study (MoBa). *International journal of epidemiology 45*, 2 (2016), 382–388. doi:10.1093/ije/dyw029. 1

[Mun15] MUNZNER T.: *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. URL: https://books.google.de/books?id=NfkYCwAAQBAJ. 1, 2

[NHG19] NUSRAT S., HARBIG T., GEHLENBORG N.: Tasks, techniques, and tools for genomic data visualization. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 781–805. 1

[PXH*18] PENG J., XUE H., HUI W., LU J., CHEN B., JIANG Q., SHANG X., WANG Y.: An online tool for measuring and visualizing phenotype similarities using hpo. *BMC genomics 19*, 6 (2018), 89–97. 1

[ukb] UK Biobank - UK Biobank — ukbiobank.ac.uk. https://www.ukbiobank.ac.uk/. [Accessed 07-09-2023]. 1