This project has received funding from
The European Union's Horizon 2020 research and innovation programme under
Grant agreement No 101017562

Finding Endometriosis using Machine Learning
FEMaLe

Call/Topic: Digital transformation in Health and Care
Type of action**:** RIA

Date: 30.06.2022

DELIVERABLE NUMBER	D4.1
DELIVERABLE TITLE	Risk classifiers
RESPONSIBLE AUTHOR	PREL (P10)

GRANT AGREEMENT No.	101017562
DOCUMENT TYPE	Report
WORKPACKAGE N. | TITLE	WP 4 | OMICS
LEAD CONTRACTOR	PREL (P10)
AUTHOR(S)	PREL (P10), AAU (P17), UOXF (P5)
PLANNED DELIVERY DATE	June 30th, 2022
ACTUAL DELIVERY DATE	July 4th, 2022
DISSEMINATION LEVEL	Public
STATUS	Completed
VERSION	Final version (1.5)
REVIEWED BY	Nemanja Todic (WBS, P14) and Tong Zhu (EQUIP, P3)

Document history

Version	Date [1]	Comment	Author	Status [2]
1.1	17-05-2022	First draft created	PREL	Drafted
1.2	27-05-2022	Second draft created, including contributions from WP4 partners.	AAU, PREL, UOXF	Drafted
1.3	15-06-2022	Third draft prepared for FEMaLe Review Panel.	PREL	Drafted
1.4	24-06-2022	Final draft created, based on FEMaLe Review Panel feedback.	PREL	Completed
1.5	01-07-2022	Final version ready for submission, quality checked by FEMaLe PMO.	AU	Validated

[1] As per the project's cloud storage or per email date if applicable.
[2] Drafted, completed or validated as per the project's cloud storage or per email date if applicable.

TABLE OF CONTENTS

Disclaimer

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

All FEMaLe Consortium members are also committed to publish accurate and up to date information and take the greatest care to do so. However, the FEMaLe Consortium members cannot accept liability for any inaccuracies or omissions, nor do they accept liability for any direct, indirect, special, consequential, or other losses or damages of any kind arising out of the use of this information.

Copyright notice

Acknowledgement

The project 'Finding Endometriosis using Machine Learning' (FEMaLe) has received funding from the European Union Horizon 2020 programme under grant number 101000640.

Citation

Be so kind as to cite this work as: Finding Endometriosis using Machine Learning, 2022: Risk classifier. FEMaLe Consortium under the supervision of the Project's Coordinator.

Legislation

Legislation H2020 Framework Programme – Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020) (OJ 347, 20.12.2013, p. 104).

H2020 Specific Programme – Council Decision 2013/743/EU of 3 December 2013 establishing the Specific Programme Implementing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p. 965).

Rules for Participation (RfP) – Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 of December 2013 laying down the rules for the participation and dissemination in Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p.81).

Financial Regulation (FR) – Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the European Union (OJ L 298, 26.10.2012, p.1).

Rules of Application (RAP) – Commission Regulation (EC, Euratom) No 1268/2012 of 29 October 2012 on the rules of application of l Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council on the financial rules applicable to the general budget of the Union (OJ L 298, 26.10.2012, p.1).

Finding Endometriosis using Machine Learning: FEMaLe

Introduction to Task 4.1 – Risk classifier

Polygenic risk scores (PRS) are popular and believed to be an important tool for precision medicine. A traditional PRS summarizes information about a person's disease risk, based on numerous genetic variants (SNPs) in their genome. Each genetic variant confers a very little increase in disease risk, but when adding them up to a composite (or polygenic) risk score, it is possible to stratify people according to a distribution of risk for a disease. In other words, they can be used to assign people to different categories of risk based on their genetic makeup. The scores are useful for people in the extreme tails of risk distribution, but less useful for most people receiving an average risk score.

In our approach we will identify combinatorial risk scores (CRS), i.e., combinations of SNP genotypes that together exhibit a non-linear epistatic effect on phenotype. We count directly how many cases-control members have these combinatorial features and how strongly they are associated with their phenotype. This not only gives a more accurate score, but also a much more personalized one. We will also benchmark the combinatorial risk score to the traditional PRS to evaluate the added value of our approach for the risk classifier.

Partners in Task 4.1 are PrecisionLife (PREL), Aalborg University (AAU), and Oxford University (UOXF).

Development of platform for discovery, replication, and risk scoring

A major task of WP4 is to develop a platform for discovery and replication of specific combinations of SNP genotypes associated with endometriosis. The platform is important not just for the current Task 4.1, but also for the upcoming Task 4.2 (stratification of subgroups of patients with endometriosis), Task 4.3 (validation and refinement of endometriosis subtypes in independent cohorts) and Task 4.4 (clinical decision support tools, based on the signature of the individual patient).

In Task 4.1 we use a dataset from UK Biobank (UKBB) for discovery of genotype combinations associated with endometriosis and an independent dataset from Copenhagen Biobank (CHB) and Danish Blood Donor Studies (DBDS) for replication of the combinations and for testing the risk profiles of the individual samples. The major workflow is sketched in the figure below.

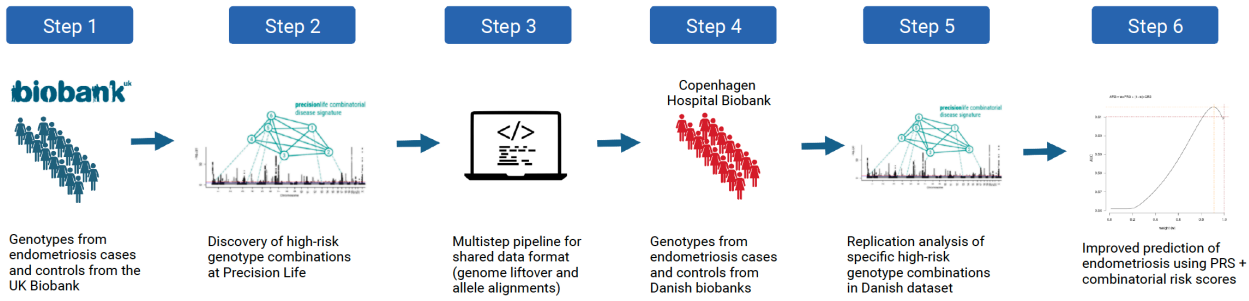

Figure 1. Overview of the Task 4.1 workflow.

The development process is based on interactive prototyping in close collaboration between the partners in WP4. PREL is responsible for the discovery process (using the patented MARKERS technology for combinatorial analytics), while the replication and risk scoring pipeline (AAU and UOXF) is independent of the discovery process in terms of software and data. Hence, we ensure the quality and independence of replication testing, while we also work on a shared data format for genotype combinations and case and controls definitions.