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: 25.01.2023

DELIVERABLE NUMBER	D4.2
DELIVERABLE TITLE	Biological markers
RESPONSIBLE AUTHOR	AAU

GRANT AGREEMENT No.	101017562
DOCUMENT TYPE	Report
WORKPACKAGE N. | TITLE	WP 4 | OMICS
LEAD CONTRACTOR	AAU
AUTHOR(S)	PREL, AAU, UOXF
PLANNED DELIVERY DATE	December 31st, 2022
ACTUAL DELIVERY DATE	January 25th, 2023
DISSEMINATION LEVEL	Confidential
STATUS	Reviewed and quality checked
VERSION	D1.5
REVIEWED BY	FEMaLe PMO

Document history

Version	Date [1]	Comment	Author	Status [2]
1.1	21-11-2022	First draft created	UOXF, AAU	Drafted
1.2	19-12-2022	Second draft created, including contributions from WP4 partners.	AAU, PREL, UOXF	Drafted
1.3	09-01-2023	Third draft prepared for FEMaLe Review Panel.	AAU, PREL, UOXF	Drafted
1.4	23-01-2023	Final draft created, based on FEMaLe Review Panel feedback.	AAU, PREL, UOXF	Completed
1.5	25-01-2023	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

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, 2023: Biological markers (Risk variants and biomarkers that define biological, endometriosis subtypes) 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).

Euratom Research and Training Programme (2014-2018) – Council Regulation (Euratom) No 1314/2013 of 16 December 2013 on the Research and Training Programme of the European Atomic Energy Community (2014-2018) complementing the Horizon 2020 – The Framework Programme for Research and Innovation (OJ L 347, 20.12.2013, p. 948).

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 4.2 – Identifying endometriosis subtypes

In Task 4.1 we developed a platform for discovery and replication of specific genotype combinations associated with endometriosis. We used the UK Biobank (UKBB) for initial discovery of genotype combinations and replication was performed within the Copenhagen Hospital Biobank/Danish Blood Donor Studies (CHB/DBDS). The replicated genotype combinations were used to identify potential genetically defined subtypes of endometriosis patients.

The aims for Task 4.2 were to investigate whether the potential genetic subtypes of endometriosis patients identified in Task 4.1 were associated with particular clinical phenotypes within the CHB/DBDS, and to initiate the investigation of circulating plasma proteins. The clinical phenotypes of the genetic subtypes were explored in 4,165 endometriosis patients from CHB/DBDS. The plasma proteins were measured with the Meso-Scale human Biomarker 54-Plex Kit in two sub-cohorts of the DBDS; a case-control cohort with 93 endometriosis cases and 4.996 female controls and a cases-only cohort consisting of 635 endometriosis cases. Additionally, we jointly led the largest endometriosis GWAS meta-analysis to date in collaboration with the International Endometriosis Genomic Consortium, in which 42 genome-wide significant endometriosis variants were identified and explored in subsequent analyses, to understand if they are driven by certain surgical or clinical features of endometriosis (Rahmioglu et al., Nature Genetics, in press); these results in turn fed into our subtype analyses under Task 4.2.

The partners in Task 4.2 are PrecisionLife (PREL), Aalborg University (AAU), and Oxford University (UOXF).

Description of data and methods

The Danish Blood Donor study (DBDS)

The DBDS is a large prospective cohort of blood donors aimed at identifying predictors of disease among healthy donors. As part of this cohort the DBDS Genomic Cohort have genotyped single nucleotide polymorphisms (SNPs) of 110,000 donors. When consenting to enrolment, participants consent for collection and analysis of blood samples for research purposes (often several collected over time).

Copenhagen Hospital Biobank (CHB)

CHB is a cohort of more than 400,000 genotyped individuals, built using the infrastructure at the Blood Banking facilities in the Capital Region of Denmark. It is based on leftover EDTA whole blood from samples drawn for blood type testing or red cell antibody screening in hospitalised patients.

Combined Danish cohort

All individuals in the two biobanks are linked to Danish registers using the Civil Registration System number, which is a 10-digit personal identification number unique for all Danish residents. To define cases and controls, we used ICD8 and ICD10 codes from registry data. This resulted in a cohort of 4,165 endometriosis cases and 48,841 female controls with genetic data available. The identified genetic endometriosis subtypes (Task 4.1) were used to test for association with (1) the clinical subtypes of endometriosis, and (2) common endometriosis disease comorbidity.

Re-discovery and replication of high-risk genotype combinations

Two important changes were made regarding discovery and replication of high-risk genotype combinations identified in Task 4.1.

First, the genetic dataset used for discovery was altered to include 42 recently discovered significant endometriosis risk variants from our meta-GWAS (Rahmioglu et al., in press), and the initial discovery analysis was repeated with the new dataset.

Second, the replication strategy was altered to allow population-specific genotype distribution. In Task 4.1 the replication of high-risk genotype combinations CHB/DBDS was based on the strategy that the genotype constellations should be **exactly** the same as those identified in UKBB in the discovery analysis. In Task 4.2 we allow the association signal to be spread to other genotype constellations than the specific genotype-combination identified in UKBB, but for the same SNP-combination. This provides more statistical power for identifying replication of the underlying biological signal.

Conclusion

In conclusion, we have replicated in an independent dataset the genotype combinations identified in UKBB and identified 15 genetic subtypes of endometriosis. Nine of the genetic subtypes were associated with particular surgical subtypes of endometriosis and some with particular co-morbidities. Six of the 42 genome-wide significant endometriosis risk variants were included in the genotype combinations and five of the 15 genetic subtypes. In addition, we have identified protein markers associated with particular subtypes of endometriosis.