Prospective study protocol	
Protocol number	1
Document version	5.3
Study Title	Prospective data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms.
Novel solution	New technology – DIOPTRA – for cancer screening and early detection
Organisation responsible for the study	Name, address
Local representative, if applicable	Name, address
Lead Principal Investigator (I)	Name, address
Coordinating investigator (CI)	Name, address

Confidentiality Statement

The information contained in this document is confidential and subject to any proprietary rights of the organisation responsible for the study. Any distribution, copying, or disclosure without the prior written authorisation of the organisation responsible for the study is strictly prohibited. Persons to whom the information is disclosed must know that it is confidential and that they may not further disclose it.

_____ _____

(position, name, surname) (signature)

Ethical principles Statement

The information in this document and related documents prepared for this study is in accordance with the recognised ethical principles (which have their origin in the Declaration of Helsinki) for medical research involving humans and the principles of good clinical practice, as well as with the applicable regional or national regulatory requirements and any additional requirements imposed by the EC or regulatory authority.

_____ _____

(position, name, surname) (signature)

Preconditions to Start Prospective Study Statement

The prospective study shall not begin until the required approval/favourable opinion from the EC and regulatory authority has been obtained.

_____ _____

(position, name, surname) (signature)

Approval sheet

_____ _____

(position) (name, surname)

_____ _____

(signature) (date)

The following persons are planned to be involved in the study:

Role	Organisation responsible for the study (S) / Principal Investigator (Clinical site) (I)/ Coordinating investigator (Clinical site) (CI)	Name	Title
Author			
Leadership			
Study team			
Study team			
Study team			
Study team			
Study team			
Quality			
Regulatory affair			
Medical writer			
Study Manager			

Contents

Funded by the European Union

1.1. DOCUMENT HISTORY

Revision	Date of enactment	Change author	Change description
2.0	2023-07-20	Amendment - AMD-101096649-12	Beneficiary termination: The beneficiary is deleted Region Midtjylland (RM-RRH) Addition of a new beneficiary: The new beneficiary is added Region Syddanmark (RSYD)
3.0	2023-10-25	Christos Androutsos (UOI)	Annex 2 update Annex 5 update
3.1	2023-11-07/17	Zheshen Jiang (CHUL) Christos Androutsos (UOI)	Changes in 4.1 DIOPTRA PARTICIPANT ID section. Each ID code is a 5-digit number instead of a 4-digit number Minor Annex 5 update Minor update in the Exclusion Criteria section
4.1	30/05/2024	Amendment No. 2	Beneficiary termination: The beneficiary is deleted BLOKS ZDRAVNI I SOTSIALNI GRIZHI EOOD (BLOCKS) Addition of a new beneficiary: The new beneficiary is added to Clinical Hospital Dubrava Zagreb/Croatia (KBDZ)
5.1	09/09/2024	Clinical Amendment No. 1 Zheshen Jiang (CHUL) Christos Androutsos (UOI) Christos Fotis (PAO) Toygar Occur (ARTHURS)	Consent forms have been updated: All documents have been revised to meet the latest project requirements. Information about the discovery study has been added: New details regarding the ongoing discovery study have been included to clarify its objectives and procedures. Information about the software component has been included: Additional information about the software component and mobile app has been added, explaining its functionality and significance in the project.
5.2	11/11/2024	Zheshen Jiang (CHUL)	Increased dropout rate from 30% to 50%. Increased participant number from N = 320 to N = 416. Follow-up study timing and procedures updated.
5.3	03/12/2024	Zheshen Jiang (CHUL)	Updated inclusion and exclusion criteria. Expanded study population description.

1. ABBREVIATIONS AND ACRONYMS

AE	Adverse event - any untoward medical occurrence, unintended disease or injury, or untoward clinical signs (including abnormal laboratory findings) in subjects, users or other persons, whether or not related to the investigational medical device and whether anticipated or unanticipated
AGSAVVAS	Geniko Antikarkiniko Ogkologiko Nosokomeio Athinon O Agios Savvas
AI	Artificial Intelligence
BURGOS	Fundacion Burgos Por La Investigacion De La Salud
CHUL	Centre Hospitalier Universitaire de Liege
CRC	Colorectal Cancer
CSCY	CSCY Computer Solutions Cyprus Ltd
FU	Follow up
GOC	Linac-Pet Scan Opco Limited
INTRA	Netcompany-Intrasoft SA
ISRCTN	International Standard Randomised Controlled Trial Number
KBDZ	Clinical Hospital Dubrava Zagreb/Croatia
NOVELCORE	D.Tsakalidis-G.Domalis OE
NKUA	National and Kapodistrian University of Athens
PAO	Protavio Ltd
PSD	Prospective Study Design
RSYD	Region Syddanmark
SOP	Standard operating procedure
TCR	Tecreando B.V.
UC	Use-Cases
UKCM	Univerzitetni Klinicni Center Maribor

UOI	Panepistimio Ioanninon

2. SYNOPSIS

Title	Prospective data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms.
DIOPTRA screening system	AI-based solution for CRC early diagnosis and screening. The main means will be biological sampling and analysis using in vitro diagnostics.
Purpose	This study aims to refine and validate the clinical use of the DIOPTRA screening system.
Prospective study design (PSD)	Prospective, cohort, multi-centre study.
PSD Primary objective	To validate the diagnostic sensitivity and specificity for CRC detection of the DIOPTRA screening system using clinical diagnosis as reference (colonoscopy).
PSD Secondary objectives	Secondary objectives include: 1. Validation of the clinical performance of the DIOPTRA screening system for detecting advanced adenomas. 2. Validation of the clinical performance of the DIOPTRA screening system for the detection of CRC in the sub-population without a prior history of malignancy. 3. Refinement of the DIOPTRA screening system. 4. Evaluation of the effectiveness of behavioural suggestions to reduce CRC risk. 5. Assessment of cost-effectiveness of DIOPTRA system.
PSD Primary endpoint	Acceptable diagnostic specificity and sensitivity for CRC detection and healthy and non-advanced adenoma groups, respectively.
PSD Secondary endpoints	Secondary endpoints include: 1. Acceptable diagnostic sensitivity for the detection of advanced adenomas. 2. Acceptable diagnostic sensitivity for CRC and advanced adenoma detection and specificity for detecting healthy and non-advanced adenoma groups in the sub-population without a prior history of malignancy. 3. Improvement of the performance metrics of the DIOPTRA screening system using the prospective data for refinement. 4. Statistically significant differences in risk factors and protein biomarker concentrations for individuals who have implemented the behavioural suggestions. 5. Improvement in the estimated efficiency of resource allocation of Significant reduction of the estimated DIOPTRA screening system costs compared to screening colonoscopy.
Duration of the study	35 months
Duration of study follow-up	1 year after follow-up enrollment.
Subject population	Individuals who visit the clinical sites for a colonoscopy.

Number of subjects	At least 1612 participants are estimated to be recruited in 8 clinical sites (KBDZ, CHUL, RSYD, UKCM, BURGOS, NKUA, GOC, AGSAVVAS) Based on sample size calculations, at least N=403 participants from each group are required to evaluate the primary and secondary endpoints of the study (A total of 1612 participants). Given the low CRC incidence rate, a much larger number is expected to participate in the study until the required numbers are recruited.
Number of Sites	8 clinical sites: 1) KBDZ 2) CHUL 3) RSYD 4) UKCM 5) BURGOS 6) NKUA 7) GOC 8) AGSAVVAS
Prospective Study Procedures	The main study procedures (observational) are the following: 1. Enrollment of participants in the study once written informed consent is obtained and subject eligibility is confirmed. 2. Blood sample collection (serum & plasma). 3. Colonoscopy & clinical diagnosis according to each clinical site's standards. 4. Collection of DIOPTRA data. 5. End of study. During enrollment, participants will be given the option to participate in the DIOPTRA follow-up study. The procedures of the follow-up study are: 1. Enrollment in the follow-up study after subject eligibility is confirmed. 2. Download the DIOPTRA mobile app to implement steps 3-5 below. 3. Answer questionnaire. 4. Receive behavioural suggestions. 5. Periodic data update. 6. Follow-up blood collection and risk assessment. 7. End of follow-up study
Study financing	This study is part of the DIOPTRA European Project, funded under Grant Agreement N° 101096649 by Horizon Europe's research and innovation program.
Person paying compensation for costs and time incurred in participating in the study, procedure, and conditions for calculation and payment of compensation	No compensation is provided*. *The compensation for the 1 (one) year FU of the validation study will depend on each site's policy. As participants will be called back for blood sampling and re-assessment as part of a non-prescribed visit, certain sites may require that compensation be provided for the travel to and back from the hospital.
ISRCTN registration	DIOPTRA's prospective study was registered under number 15583857 on 26 October 2023 (https://doi.org/10.1186/ISRCTN15583857)

3. BACKGROUND & RATIONALE

BACKGROUND

Incidence & Survival Rates. Colorectal cancer [1], [2] (CRC) is the third most common cancer in men and the second in women, accounting for 10% of all tumours worldwide. It ranks second in cancer-related deaths with 9.4%, only below lung cancer. About 1.9 million new cases were diagnosed in 2020, translating into 0.9 million deaths, while incidence is projected to rise significantly over the next decade, with 3.2 million new diagnoses annually by 2040. In affected EU individuals, 5-year survival ranges from 28.5% to 57% in men and 30.9% to 60% in women, with pooled estimates in 23 countries of 46.8% and 48.4%, respectively. Moreover, CRC is among the five most likely to metastasise cancers. Upon initial diagnosis, 22% of cases are metastatic, while about 70% of patients will eventually develop metastatic relapse [3].

Existing Standard & Screening Impact. Screening methods consisting of endoscopic tests (e.g., colonoscopy) and non-invasive alternatives such as the faecal immunochemical test (FIT) have been put into action [1]. Studies have compared mortality rates for symptom-detected vs. screening-detected CRC, stating the considerable impact of screening via quantified reduction estimates surpassing 30% for screening-based detections [5]. Notably, the 5-year survival rate can reach 90% for stage I diagnosis, being less than 15% for advanced stages [6]. Therefore, routine screening is vital for reducing mortality and declining incidence rates since CRC is now considered a highly preventable disease with a considerably wide temporal development window [7]. Namely, the transitional path from normal mucosa to pre-malignant growth and then to malignant lesion might spread over 15 to 20 years, with scientists seeking means for earlier, cost-effective, and less taxing detection of premalignant states.

Pressing Conditions. In determining the CRC risk status, factors such as age, BMI, diet, smoking habits, and family history [4] have been pinpointed by researchers and clinicians alike. Despite the long-assumed CRC preventability based on modifiable risk factors, awareness and knowledge exploitation remain extremely low. Overall, taxing procedures, citizen reluctance, poor awareness, and screening accessibility are hindering participation, forcing researchers into the survey of accessible, non-invasive biomarkers that bear the potential to render cancer screening less burdensome and more accessible to citizens.

Liquid Biopsy CRC Biomarkers. Liquid biopsy is a promising new tool for noninvasive, quick, and safe assessment [5]. Blood-derived proteins constitute the most cost-effective solution among all liquid biopsy products, judging by resources, sensitivity, and research maturity. On this premise, a vast protein

pool has been tested, albeit evidence lacks comparative validation and perplexing standardisation margins. Research must highlight a small biomarker subset that can be feasibly exploited for population-based screening and sustainably covered by health insurance bodies.

Artificial Intelligence for Cancer Screening. AI has been widely employed in biomarker evaluation, from drug development to pathology and oncology [11]. However, despite the AI advances in CRC risk and progression assessment, the medical community is still sceptical and reluctant to trust the outcomes of machine learning. This is mainly due to the depth of most neural network approaches and the confusing architecture, which are regarded as "black boxes" [12]. Explainable artificial intelligence (XAI) is gradually becoming a prerequisite for clinicians and policymakers seeking to instil accountability and medical transparency into AI-assisted decisions for launching trustworthy clinical applications [13].

Risk Factor Analysis. Numerous studies have investigated the association of CRC incidence with demographic, behavioural, and environmental risk factors, including age, sex, and lifestyle. Age is the main factor assessed by current guidelines, forming at-risk groups for recommended screening [14]. Clinical practice has shown that these thresholds are gradually decreasing, a fact under study by the medical community. Several lifestyle-related factors have been identified and modifiable through suitable behavioural screening and personalised interventions.

RATIONALE FOR THE PROSPECTIVE STUDY

DIOPTRA aims for an accessible and less taxing screening to attain a more comprehensive population outreach by exploiting blood-based biomarkers. Although several researchers have tried to assess this, the limitations in the number of participants and number of proteins studied hinder a generalised framework for early CRC screening and prevention. The same applies to AI-enabled CRC risk assessment, where clinical validation of established systems [2] is absent. DIOPTRA focuses on fulfilling the role of biomarker identification and risk factor stratification via validation on 8 different sites, utilising a large number of patients/ healthy citizens. As such, the evidence produced will not only be based on expert opinion but also vigorous validation procedures on the retrospective and prospective data (level of evidence B). The study's retrospective part hypothesised that predictive variables are associated with the risk of developing CRC. Based on this study, risk factors will be identified to investigate their association with CRC and predict the early risk of CRC. Data from electronic health records will be used and analysed to isolate variables defined as risk factors based on four groups. Various methodologies, including statistical analysis and machine learning techniques, will be used to investigate the impact of each factor on CRC. More importantly, by employing cutting-edge *in vitro* protein analysis in (paired with the blood samples collected) biopsies, the molecular mechanism of CRC development will be uncovered, providing additional evidence needed and thus establishing a robust and efficient framework for early screening (level of evidence A). The mobile application will serve as

an essential tool, allowing participants to engage with the study seamlessly, access personalised assessments, and receive tailored suggestions.

4. STUDY DESIGN

The study will be a prospective, cohort, multi-centre study with a partial follow-up of one year. Changes to the recruitment process throughout the duration of the project are not envisaged. An equal sample size will be required for all 4 groups: healthy, non-advanced adenomas, advanced adenomas and CRC cases. Only the first two groups will be enrolled in the follow-up study. Initial data obtained will be used for the algorithm training, followed by pilot validation.

STUDY HYPOTHESIS

The main study hypothesis is that the DIOPTRA screening system has adequate clinical performance to diagnose CRC and advanced adenomas early. An additional hypothesis is that the DIOPTRA system can accurately characterise an individual's risk of developing CRC. Finally, we hypothesise that the DIOPTRA behavioural suggestions, when applied, can significantly lower the risk of developing CRC. To evaluate these hypotheses, we will use multiplex protein biomarker measurements and demographic, behavioural, and clinical data from participants in the DIOPTRA study groups to test and refine the DIOPTRA AI models.

TYPE OF INTERVENTION

Biological samples will be collected via a minimally invasive method. According to each clinical site’s standards and pre-existing practice, enrolled individuals will undergo a screening colonoscopy, while blood will be drawn (prior to the colonoscopy) for the purposes of the study. All biological data will be used for *in vitro* protein-based analysis, allowing the construction of preliminary decision algorithms and AI analysis models.

SITES OF THE PROSPECTIVE STUDY

Research center No 1: Univerzitetni klinicni center Maribor
Address: LJUBLJANSKA ULICA 5, 2000 Maribor, Slovenia
Tel: 02/321 1000; e-mail: gpisarna@ukc-mb.si
Title of the department (s): Department of Oncology
Tel: 02/321 1974; e-mail:.
Principal investigator: Maja Ravnik, MD
Tel: 02/321 1975; e-mail: maja.ravnik@ukc-mb.s

Research center No 2: Hospital Universitario de Burgos
Address: Avenida Islas Baleares 3, Bloque F, Planta 0, CP:09006, Burgos, España
Tel: 947256533- ext. 36078.; e-mail: ralcaraz@hubu.es
Title of the department (s): Department of Oncology
Tel: 947256533- ext. 36078; e-mail:
Principal investigator: Enrique Lastra Aras
Tel: 947256533- ext. 36078; e-mail: elastra@saludcastillayleon.es

Research center No 3: Centre Hospitalier Universitaire de Liège
Address: Avenue de l'hopital, 1 CP: 4000 Liège, Belgium
Tel: +32(0)4.323.00.00; e-mail: info@chuliege.be
Title of the department (s): Department of Gastroenterology, Hepatology and Digestive Oncology
Tel: +32(0)4.323.72.56; e-mail: gastroentero@uliege.be
Principal investigator: Dr. Catherine Loly
Tel: +32(0)4.323.72.56; e-mail: gastroentero@uliege.be

Research center No4: German Oncology Center
Address: 1 Nikis Avenue, 4108 Agios Athanasios, Limassol, Cyprus
Tel: +35725208000; e-mail: info@goc.com.cy
Title of the department (s): Research and Innovation
Tel: +35725208159; e-mail: kristis.vevis@goc.com.cy
Principal investigator: Dimitrios Giannakakis
Tel: +35725208000; e-mail: dimitrios.giannakakis@goc.com.cy

Research center No 5: Agios Savvas
Address: Leof. Alexandras 171, Athina 115 22
Tel: 21 0640 9000; e-mail: info@agsavvas-hosp.gr
Title of the department (s): N/A
Tel: N/A; e-mail: N/A
Principal investigator:Dimitrios Kypraios, MD & Charilaos Maragkos, MD
Tel: 21 0640 9000; e-mail: poiotita@agsavvas-hosp.gr

Research center No 6: Region Syddanmark
Address: Beridderbakken 4, 7100 Vejle
Tel: +45 79405000; e-mail: N/A
Title of the department (s): Department of Data, Innovation and Research
Tel: N/A; e-mail: N/A .
Principal investigator:Sisse Helle Njor
Tel: +45 79405000; e-mail: sisse.helle.njor@rsyd.dk

Research center No7: ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON (NKUA)
Tel: N/A.; e-mail: N/A
Title of the department (s): Hepatogastroenterology Unit, 2nd Department of Internal Medicine and Research Unit, Attikon University General Hospital
Address: 1 RIMINI STR, Chaidari, Greece, 12462
Tel: +302105832090; e-mail: gastr@attikonhospital.gr .
Principal investigator: Dr. Ioannis Papanikolaou
Tel: +302105832090; e-mail: ispapn@med.uoa.gr.

Research center No8: Clinical Hospital Dubrava, Zagreb/Croatia, University of Zagreb Faculty of Pharmacy and Biochemistry
Address: Avenija Gojka Šuška 6, Zagreb, Croatia, 10000
Tel: +385 1 290 2550; e-mail: N/A
Title of the department (s): Department of Gastroenterology, Hepatology and Clinical Nutrition
Tel: +385 1 290 2550; e-mail: N/A .

Principal investigator: Dr. Mario Tadic
Tel: + 38598844606; e-mail: mtadic1@gmail.com

5. OBJECTIVES

This study aims to refine and validate the clinical use of the DIOPTRA screening system.

Primary objective: to validate the diagnostic sensitivity and specificity for CRC detection of the DIOPTRA screening system using clinical diagnosis as reference (colonoscopy).

Secondary objectives:

1. Validation of the clinical performance of the DIOPTRA screening system for detecting advanced adenomas.
2. Validation of the clinical performance of the DIOPTRA screening system for detecting CRC in the sub-population without a prior history of malignancy or concurrent malignancy*.
3. Refinement of the DIOPTRA screening system.
4. Evaluation of the effectiveness of behavioural suggestions to reduce CRC risk.
5. Assessment of cost-effectiveness of the DIOPTRA system.

*Prior history of malignancy or concurrent malignancy according to the inclusion criteria:
Prior history of malignancy: other than CRC, if all treatment of that malignancy is completed at least 2 years before registration and the patient has no evidence of disease.
Concurrent malignancy: Concurrent, clinically stable malignancy, other than CRC, without previous treatment, that does not require tumor-directed treatment.

Funded by the European Union

6. ENDPOINTS

Primary endpoint. Acceptable diagnostic specificity and sensitivity for CRC detection and healthy and non-advanced adenoma groups, respectively.

Secondary endpoints include:

1. Acceptable diagnostic sensitivity for the detection of advanced adenomas.
2. Acceptable diagnostic sensitivity for CRC and advanced adenoma detection and specificity for detecting healthy and non-advanced adenoma groups in the sub-population without a prior history of malignancy or concurrent malignancy*.
3. Improvement of the performance metrics of the DIOPTRA screening system using the prospective data for refinement.
4. Statistically significant differences in risk factors and protein biomarker concentrations for individuals who have implemented the behavioural suggestions.
5. Improvement in the estimated efficiency of resources allocation of DIOPTRA screening system costs compared to screening colonoscopy.

*Prior history of malignancy or concurrent malignancy according to the inclusion criteria:
Prior history of malignancy: other than CRC, if all treatment of that malignancy is completed at least 2 years before registration and the patient has no evidence of disease.
Concurrent malignancy: Concurrent, clinically stable malignancy, other than CRC, without previous treatment, that does not require tumor-directed treatment.

Funded by the European Union

7. STUDY POPULATION

The prospective study will cover at least 1600 participants to be recruited across all the study's 8 clinical sites (KBDZ, CHUL, RSYD, UKCM, BURGOS, NKUA, GOC, AGSAVVAS). The study population will cover participants who visit the clinical sites for a colonoscopy. After a high-quality colonoscopy is conducted, with complete removal of all detected neoplastic lesions, participants will be split into the following groups following the histopathological analysis of index lesions identified during colonoscopy:

- Healthy.
- Non-advanced adenoma (NAA).
- Advanced adenoma (AA).
- Colorectal cancer (CRC).

Because of the large heterogeneity of these groups, participants in each category will be further characterised in more detail, as described in the table below. In particular, participants of the healthy group will be further categorised into two different subcategories, one without any neoplastic lesion detected (1.1) and the other with hyperplasic polyps < 10mm (1.2). Participants in the NAA group will be characterised with a, b or both depending on the presence of adenoma and/or SSP/SSA and further categorised based on their number. Finally, participants in the AA group will be characterised with subcategories a to f according to the histopathology report and the number of adenomas detected.

Table 1. DIOPTRA group and subgroup definitions

<table>
<tr><th colspan="2">DIOPTRA group</th><th colspan="3">Subgroups</th></tr>
<tr><td rowspan="2">1</td><td rowspan="2">Heathy</td><td colspan="2">1.1</td><td>No findings</td></tr>
<tr><td colspan="2">1.2</td><td>Hyperplastic polyps, <10 mm</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">Non-Advanced Adenoma</td><td rowspan="4">2a NAA <10mm, tubular, no HGD
AND/OR
2b SSP/SSA <10mm, no dysplasia</td><td>2 a.1</td><td>1-2 NAA</td></tr>
<tr><td>2 a.2</td><td>3-4 NAA</td></tr>
<tr><td>2 b.1</td><td>1-2 SSP/SSA <10mm, no dysplasia</td></tr>
<tr><td>2 b.2</td><td>≥3 SSP/SSA <10mm, no dysplasia</td></tr>
<tr><td>3</td><td>Advanced</td><td colspan="2">a</td><td>≥5 Adenomas</td></tr>
</table>

	Adenoma	b	Adenoma ≥ 10 mm
		c	Villous growth pattern >25%
		d	High-grade dysplasia (HGD)
		e	Cancer in situ (CIS)
		f	SSP/SSA ≥ 10 mm and/or dysplasia
4	CRC	4	By clinical stage

This study will take into account gender distribution in the incidence of CRC, including rates similar to those of male and female participants. However, males are 25% more prone to develop CRC than females, which could lead to a greater number of male participants in the study.

Inclusion criteria for prospective data collection and pilot evaluation:

- Any indication for total colonoscopy (including routine screening and presence of symptoms/ FIT positive).
- Previously treated malignancy, other than CRC, if all treatment of that malignancy is completed at least 2 years before registration and the patient has no evidence of disease.
- Concurrent, clinically stable malignancy, other than CRC, without previous treatment, that does not require tumor-directed treatment
- Age between 18-80 years at the moment of recruitment (see above)
- Absence of significant comorbidities (ASA IV)
- Ability to provide valid (written informed) consent
- Boston-Bowel-Preparation-Scale (BBPS) left/transverse/right colon score ≥2, total score ≥6

Inclusion criteria for the follow-up study patients who will use the DIOPTRA mobile application:

- Presenting the 4 inclusion criteria here above.
- Patients willing to use the DIOPTRA application regularly.
- Level of digital literacy allows managing mobile terminals (smartphones, smartphone apps, tablets).
- Good internet connection coverage at home.
- Availability of a smartphone/ tablet (to use the app).
- Belonging to the healthy or non-advanced adenoma groups

Funded by the European Union

Exclusion criteria.

Persons belonging to the vulnerable group will not be included in the clinical study.

Other exclusion criteria for the prospective study:

- Age under 18 y/o or above 80 y/o
- Comorbidities ASA IV
- Any condition (e.g. prior major abdominal surgery, prior abdominal or pelvic radiation therapy) that in the endoscopist's opinion could predispose to incomplete colonoscopy
- History of colectomy for reasons other than CRC
- Malignancy other than CRC, completely treated in the last 2 years prior to the registration or with clinical evidence of disease
- Concurrent malignancy, other than CRC, clinically unstable or requiring tumor-directed treatment
- Inflammatory bowel diseases
- Polyposis syndrome
- Pregnancy or suspicion of pregnancy
- Colorectal cancer history
- BBPS left/transverse/right colon score <2, total score <6
- Not able to understand the study and provide valid consent

Exclusion criteria for the follow-up study:

- Classification in the CRC or advanced adenoma groups
- Non-availability of a smartphone/tablet or inability to use a mobile app (e.g., due to low digital literacy)

Funded by the European Union

8. SAMPLE SIZE

To evaluate the study's endpoints (diagnostic sensitivity and specificity of DIOPTRA), the exact binomial test will be used with the NULL hypothesis. $H_0: p \leqq p_0$, where p_0 is the pre-specified lower bound of the endpoint and p is the observed endpoint in the sample. The pre-specified lower bounds of the endpoints were selected based on the decision memo (CAG-00454N) from the Centers for Medicare & Medicaid Services (CMS) to cover a blood-based biomarker test as an appropriate colorectal cancer screening test.

In terms of the primary endpoint (diagnostic sensitivity for CRC detection), with the following assumptions:

- Required power = 0.8
- Confidence level = 0.05
- Lower bound of sensitivity = 0.74
- DIOPTRA sensitivity hypothesis = 0.8
- Safety factor = 1.33 (taking into account the removal of participants due to poor sample/data quality and small deviations in the sensitivity hypothesis)

The required sample size is N = 403 participants in the CRC group. Using the same sample size for each of the healthy and non-advanced adenoma groups, along with the following assumptions:

- N = 403
- Confidence level = 0.05
- Specificity lower bound = 0.9
- DIOPTRA specificity hypothesis = 0.94

The statistical power of the exact binomial test for specificity is 0.91 (per group). The power to reject the NULL hypothesis is satisfactory.

Assuming the same sample size N = 403 for the advanced adenoma group and the following assumptions:

- N = 403
- Confidence level = 0.05
- Advanced adenoma sensitivity lower bound = 0.42
- DIOPTRA advanced adenoma sensitivity hypothesis = 0.5

The statistical power of the exact binomial test for the sensitivity of advanced adenoma detection is 0.94. The power to reject the NULL hypothesis is satisfactory.

The confidence intervals for the diagnostic performance metrics of the study, using the calculated sample sizes, are shown in the table below.

Funded by the European Union

Table 2. Diagnostic performance confidence intervals

Endpoint	DIOPTRA hypothesis	95% CI
Sensitivity for CRC detection	0.8	[0.76,0.84]
Diagnostic specificity (healthy & non-advanced adenomas)	0.94	[0.91,0.96]
Sensitivity for advanced adenoma detection	0.5	[0.45,0.55]

Participants who agree to enrol in the follow-up DIOPTRA study will be split into two groups. The Case group will receive behavioural suggestions via the app to lower their CRC risk score, while the Control group will not. Each of these two groups will be subdivided into two groups, which will contain patients with healthy and non-advanced adenomas, will be included. The multiplex protein biomarker readouts at the initial visit and follow-up will be compared using the t-test for paired samples. The NULL hypothesis of the statistical test is that for each DIOPTRA study group, there is no difference in the mean of protein biomarkers measured from blood samples at the initial and follow-up stages. With the following assumptions:

- Normal distribution of biomarker readouts,
- Cohen's d = 0.4 (moderate effect size),
- Significance level = 0.05,
- Power = 0.8,
- Two-sided comparison,
- Dropout rate = 50%.

The required sample size to reject the NULL hypothesis is N = 104 (N = 52 before taking the dropout rate into account) for each DIOPTRA study group in the Case and Control groups. In total, N = 416 participants are required.

Additionally, for the comparison of the CRC risk score (expressed as a percentage) between the initial and follow-up stages, using the following assumptions:

- Effect size h = 0.4 (moderate to large),
- Significance level = 0.05.

The calculated sample size (N = 52) results in a statistical power of 0.89 to detect moderate to large differences in the CRC risk score after implementing DIOPTRA behavioural modification suggestions.

Funded by the European Union

9. STUDY PROCEDURES

9.1 OVERVIEW

For the main observational study, the study procedures are the following:
Enrollment of participants in the study will be done once written informed consent is obtained and subject eligibility is confirmed.

1. Blood sample collection (serum & plasma).
2. Colonoscopy & clinical diagnosis according to each clinical site's standards.
3. Collection of DIOPTRA data.
4. End of study.

During enrollment, participants will be given the option to be potentially contacted to participate in the DIOPTRA follow-up study. The procedures of the follow-up study are as follows:

1. Enrollment in the follow-up study after subject eligibility is confirmed.
2. Download the DIOPTRA mobile app to implement steps 3-5 below.
3. Answer the questionnaire.
4. Receive behavioural suggestions.
5. Periodic data update.
6. Follow-up blood collection and risk assessment.
7. End of follow-up study.

9.2 ENROLLMENT

Individuals who visit the hospital sites with an invitation for a total colonoscopy, including routine screening or due to symptoms, will be invited to participate in the DIOPTRA study. Subjects are considered enrolled participants once written informed consent is obtained, and subject eligibility is confirmed according to the inclusion and exclusion criteria.

9.3 BLOOD SAMPLE COLLECTION

Approximately 20 mL of peripheral blood will be collected from each participant. The blood sample collection, management, and storage will be performed according to the SOP: "Sample Collection & Management" provided by PAO (Annex No. 3).

Funded by the European Union

9.4 COLONOSCOPY AND DIAGNOSIS

Each participant will undergo a colonoscopy procedure following the blood sample collection. The colonoscopy should be completed within 30 days of enrollment. The procedure and preparation will follow each site's clinical standards. During the procedure, the study personnel must fill out the "Colonoscopy and Sample Collection Case Form" provided in Annex No.4. This form contains information regarding the quality of the colonoscopy (preparation and procedure) and the collected blood samples. Participants with inadequate bowel preparation (i.e. Boston Bowel Preparation Scale overall score <6 or score in any colon segment <2) or incomplete colonoscopy due to technical factors (including but not limited to redundant or tortuous colon, marked diverticular disease, fixation of colonic loops, adhesions due to previous surgery) or due to intolerance, resulting in an incomplete procedure, will be excluded from the study. However, participants in whom colonoscopy cannot be completed due to obstructive colorectal cancer will be included in the study. Index lesions will be biopsied and sent for diagnostic analysis during the colonoscopy according to each site's clinical standards. The study personnel will gather the diagnostic data from the biopsies, follow the analysis, and match it to the participant's records and forms. The diagnostic results will be assigned to the participants in the DIOPTRA study groups.

9.5 COLLECTION OF DIOPTRA DATA

Demographic, lifestyle and behavioural data corresponding to potential risk factors for CRC will be collected during the study via the DIOPTRA behavioural questionnaire. Additionally, medical data, personal and family history, and symptoms will be collected during the study via the "Medical Information / History Case Form" (Annex No 5). The study personnel will upload all collected DIOPTRA data during the study to the DIOPTRA prospective platform.

9.6 END OF STUDY

Participants will be considered completed from the main observational study when they have provided all DIOPTRA data and completed their colonoscopy procedure or at the point of subject withdrawal.
The study will be initiated on M1 after the clinical study preparation and ethics approvals. Table 3 provides an overview of the study schedule, followed by all clinical sites.

Funded by the European Union

Table 3. Overview of the study schedule

Description	Timing
Study initiation	After ethical approval
Enrollment of participants in the study	Following informed consent and eligibility confirmation
Blood sample collection (serum & plasma)	Following enrollment
Colonoscopy & clinical diagnosis according to each clinical site's standards	Colonoscopy, no later than 30 days of blood draw. Diagnosis timing: approximately 3 months after the procedure
Collection of DIOPTRA data	From study initiation until the end of the study
Follow-up study	The follow-up study will be initiated at month 26 of the project
Follow-up study: 1. Participant recruitment 2. Behavioural data collection 3. Biological sample collection 4. Possible participant self-reported symptoms module integration in the mobile app 5. Data analysis	Approximately 1 year after follow-up enrolment

9.7. FOLLOW-UP STUDY TIMING AND PROCEDURES

Based on the recruitment rate estimated on each clinical site during the retrospective study, the maximum duration of follow-up study recruitment has been estimated to 3 months. The follow-up visit is at 1 year, considering the progression of mobile app development and the project's lifespan. (*Figure 1*)

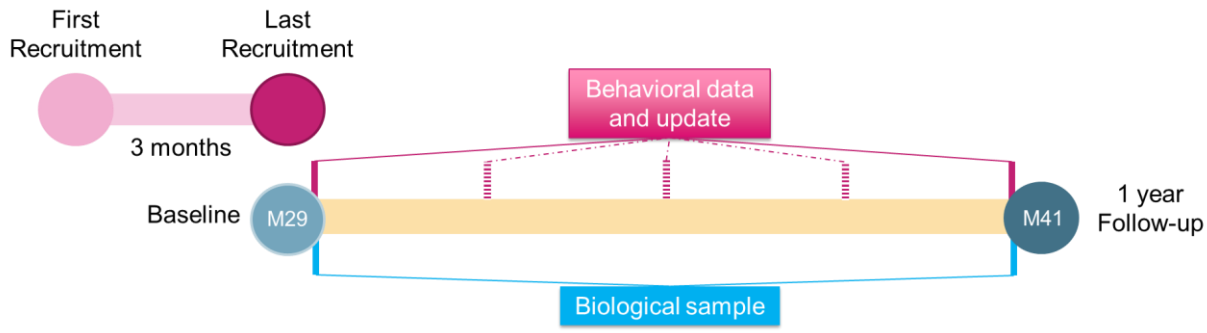

Figure 1. Follow-up study timeline

9.7.1. Participant recruitment (M26-M29)

Table 4. Participant recruitment from each clinical site

Participants/clinical site	Healthy	Non-advanced Adenomas
Case (Suggestions)	13	13
Control (No suggestions)	13	13

Prospective cohort enrollment will include participants corresponding to inclusion and exclusion criteria for the follow-up study. After the consent signature, individuals will be split into two groups. The Case group will receive behavioural modification suggestions via the DIOPTRA mobile app, while the Control group will not. In each group, participants belong to two DIOPTRA study subgroups: healthy and non-advanced adenoma. In total, 416 participants from the same 8 clinical sites will be recruited. Randomisation will be performed to ensure against bias, using appropriate randomisation methods such as block randomisation and adaptive randomisation.

9.7.2. Follow-up data collection (M29, M41 and in-between)

All participants will answer the baseline behavioural data questionnaire in the DIOPTRA mobile app at the moment of recruitment or the first moment the participant is available to answer the questionnaire after recruitment. A modified follow-up questionnaire will be answered at 1-year follow-up. Depending on the results of retrospective data analysis, adaptive questions will be updated and asked by a mobile app (more details will be generated in the DIOPTRA requirement work package). Depending on the DIOPTRA system construction progress, a participant self-reported symptoms module could be

integrated into the mobile app to allow participants to report mild symptoms that don't need medical attention.

9.7.3. Biological sample collection (M29 and M41)

Blood samples will be collected from each participant at baseline M29 and at the end of the study M41. Samples will be processed and sent to PAO for biomarker analysis. Details of sample processing can be found in Annex 2, with an additional Sample Collection Form in Annex 4.

9.7.4. End of follow-up study

The follow-up study will end 12 months after the initial recruitment at M30, as shown in Figure 1.

9.8 DATA FLOW AND DATA PROCESSING OF PROSPECTIVE AND FOLLOW-UP STUDY

The data flow and processing for the prospective and follow-up study are critical to ensuring accurate and efficient collection, analysis, and interpretation of participant data. This section outlines the systematic approach to managing the data gathered during the study, from initial data collection at clinical sites to the final analysis and integration within the DIOPTRA system. The following sections detail the key stages of data flow and processing, emphasising the interaction between various components and the centralised data management within the DIOPTRA Software.

9.8.1 DIOPTRA software components

The project's implementation includes the development and/or integration of the following software components listed in Table 5:

Table 5. DIOPTRA software components and relevant Partners

No	Component	Short Description	Responsible Partner
A	DIOPTRA Software Frontend - Anonymization Tool (EHR data)	To safeguard the confidentiality and privacy of sensitive data within the prospective study, an anonymisation tool is provided by the technical partners to empower clinical partners in the secure handling of the retrospective data. .	CSCY
B	DIOPTRA Software Frontend - Clinical Dashboard	Interface for the clinicians to manage the uploaded data.	CSCY

C	DIOPTRA Software Backend	It is responsible for data ingestion and retrieval, data curation and storage and user management.	INTRA
D	Mobile Application	This component, among other functionality (knowledge base, recommendations, etc.), collects and uploads behavioural questionnaire data from follow-up study participants.	TCR
D.a	Risk Assessment Module	Recommender system providing personalised guidance based on the participant's behavioural data.	UOI
E	Risk Factor Analysis	The Risk Factor Analysis aims to identify and evaluate the association of various risk factors with CRC and to predict early CRC risk.	UOI
F	AI Modelling	The component exploits advanced Artificial Intelligence techniques to classify individuals into four distinct categories: healthy, Non-advanced adenomas, Advanced adenomas, and Colorectal Cancer. It employs machine learning algorithms to identify significant biomarkers by utilising protein concentration data and various risk factors from the prospective study.	NOVELCORE

Annex No. 6 of this protocol describes all the components extensively, accompanied by diagrams specifying each component's outputs and communication with the other components.

9.8.2 Overall data flow

The overall data flow within the DIOPTRA project is illustrated in the following Figure 2, detailing the interaction of the components and processes. Initially, the anonymisation tool is applied at the clinical premises to provide additional data privacy protection. All data gathered at each clinical site is encoded based on the standardised data template before being uploaded to the DIOPTRA Software, where it will be securely stored. The DIOPTRA Software acts as the unique repository of all the data, metadata, and results from the analyses, ensuring centralised management and access. The project's infrastructure, hosted by GRNET, will support the development and operation of all components. Biological samples collected from participants are delivered for analysis to extract protein features. The results of these analyses are then uploaded to the DIOPTRA Software for storage. The Risk Factor Analysis component retrieves the prospective data, specifically tabular Electronic Health Record (EHR) data, to examine and identify risk factors for CRC. The findings from this analysis are also stored back in the DIOPTRA Software. The AI Modeling component will access the stored protein features and the identified risk factors from the DIOPTRA Software to predict the CRC risk. The results of these predictions are also stored within the DIOPTRA Software. The mobile application communicates with the DIOPTRA Software to retrieve a subset of the prospective dataset, specifically the behavioural data. This data serves as input for the RAM, which generates a wellness score and tailored recommendations for the participant. The generated information is then sent back to the DIOPTRA Software, where it is stored

and made available for visualisation. Clinicians can access these visualisations to assess the behavioural profile of the participant, aiding in their overall health management and intervention planning.

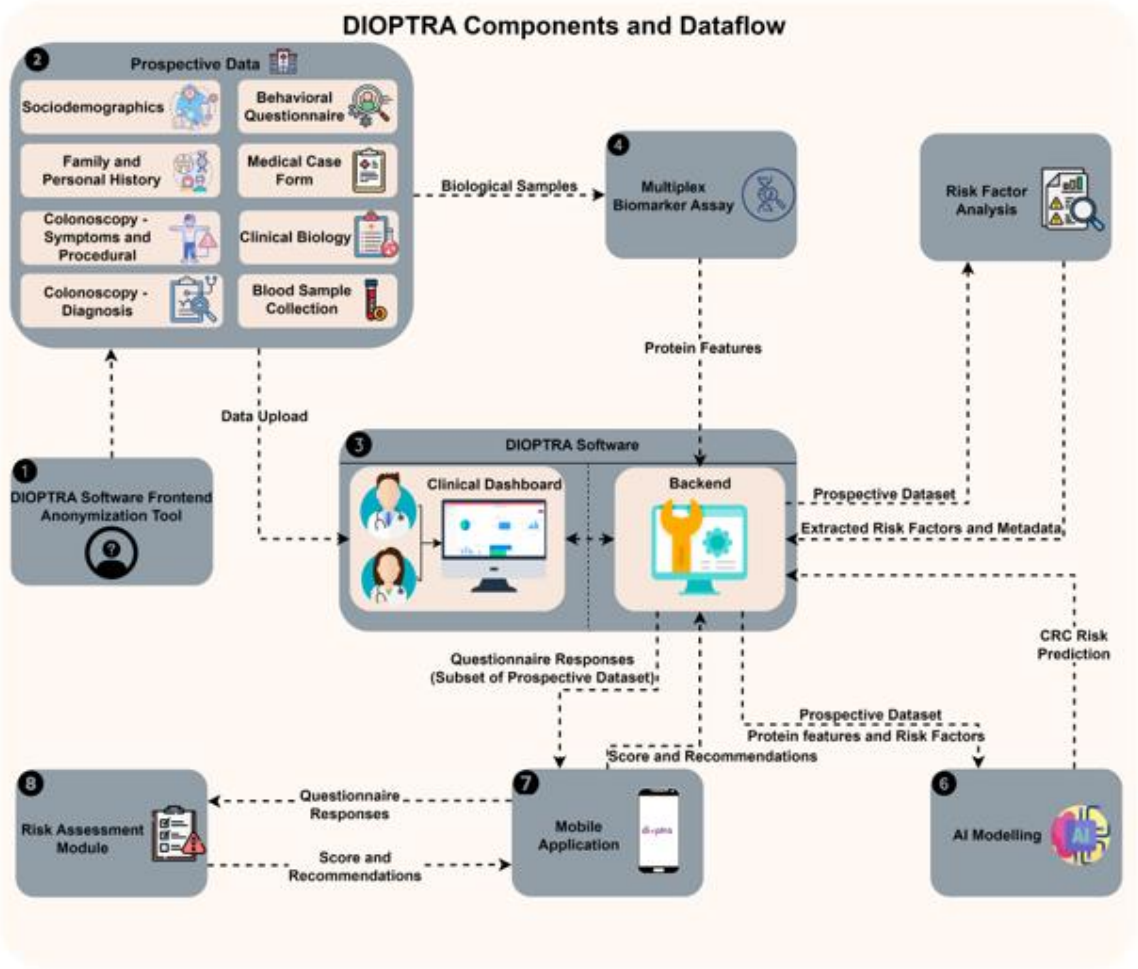

Figure 2. Components and overall dataflow

9.9 DATA ANALYSIS

Serum and plasma samples will be analysed to quantify the levels of the DIOPTRA protein biomarkers using multiplex proteomics. In terms of multiplex proteomics, the xMAP method will be used (Luminex Corp). Multiplex assays that utilise the xMAP technology rely on colour-coded microspheres (bead regions) to allow for the simultaneous detection of responses against multiple protein targets from the same sample. Each bead region is coated with an antibody that recognises and binds to a specific part of the protein. Mixtures of bead regions are used in a sandwich-type ELISA assay to provide absolute quantification of multiple proteins across the different conditions tested. These assays offer high multiplexability, sample throughput, quality of measurements, and specificity for the measurement of

the identified biomarkers in serum and plasma. The multiplex biomarker readouts, along with the various behavioural, demographic, and clinical risk factors, will be used as input to validate and refine the DIOPTRA AI screening models.

In terms of AI screening models, several methodologies will be examined. Machine Learning algorithms like logistic regression, Support Vector Machines (SVM), Naive Bayes, Decision Trees, Random Forests (RF) and Gradient Boosting (GB) can be trained on protein biomarkers data and risk factor information to classify participants as being at high risk of CRC. Feature selection (e.g. regularisation) techniques can help identify the most informative biomarkers and risk factors for efficient and accurate screening.

To evaluate the endpoints related to the clinical performance of the DIOPTRA screening system in terms of sensitivity and specificity, the confusion matrices between the reference method (colonoscopy & diagnosis) and the DIOPTRA system will be utilised. On this front, the confusion matrices (*Table 6*) for the different endpoints and models will be constructed. Diagnostic metrics will be calculated as Sensitivity = TP/(TP+FN) and Specificity = TN/(TN+FP). The exact binomial test will be utilised to compare the performance metrics to their respective lower bounds. Additionally, the confidence intervals will be calculated using the Clopper-Pearson exact method. The endpoints related to significant differences between categorical and numerical variables, i.e., the CRC risk factors and protein measurements following the behavioural suggestions, will be evaluated using 1) chi-squared test, 2) ANOVA, 3) Generalised least squares for multi-level factors, 4) T-test and other statistical methods. Finally, the improvement in performance following the refinement of the DIOPTRA models using the prospective data will be evaluated using the exact binomial test and their respective confidence intervals.

Table 6. Confusion matrix

		DIOPTRA Predictions	
	Total Population (P+N)	Positive (PP)	Negative (PN)
Reference	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

10.ETHICS

Transparency, accountability, respect for privacy, and protection of human well-being are the core ethical principles of the DIOPTRA clinical studies. Following these principles, the DIOPTRA consortium ensures that prospective clinical study participants are well-informed about the purpose and nature of the research and made aware of risks, rules, safeguards, and rights concerning their participation. For this purpose, two separate consent forms are used for different target groups of participants. Annex 1. A is used for the general prospective study where recruited individuals will only participate in the study during their colonoscopy visit to the hospital. Annex 1. B will be used for individuals recruited to the follow-up study using the DIOPTRA mobile application, with a second visit to the hospital after one year of follow-up for blood sampling.

Furthermore, the organisation responsible for the clinical study will communicate information to the participants regarding the purpose and nature of the research, potential risks, the processing of personal data and the organisational and technical safeguards implemented in the project. The participants will provide their informed, explicit consent to voluntary participation in the clinical study and the processing of their personal data through the consent forms annexed to this protocol.

Revision in Patient Information and Informed Consent Form. The organisation responsible for the study will inform the investigator whenever new information that may be relevant to the participant's confirmed participation in the study becomes available. The investigator should inform the participants or their proxy or legal guardian in a timely manner and in clear and plain language.

The organisation responsible for the study will revise the written Informed Consent Form whenever new information that may be relevant to the subject's confirmed participation in the study becomes available. The revised consent forms will be sent to the investigator for approval by the Bioethical Committee/ other regulatory authorities. After approval by the Bioethical Committee as applicable, a copy of this information must be provided to the participating subjects, and the informed consent process needs to be repeated.

The organisation responsible for the clinical study has executed a data processing agreement with the DIOPTRA consortium partners, which will process the participant's personal data to ensure that appropriate technical and organisational measures have been implemented to provide a level of security and protection of the personal data. The project data management plan, defined in Deliverable D1.1, describes the project risk management, including identification, assessment and prioritisation of risks related to data processing and explains the organisational and technical measures to eliminate or minimise and control the impact of the identified risks.

Regulatory submission. No subjects will be enrolled in the study until all necessary approvals (e.g., by the Bioethical Committee of each DIOPTRA clinical partner and/or other competent authorities) have been obtained.

11.QUALITY CONTROL PROCEDURES

Data review and processing. Before study initiation, a representative of the study consortium will review the protocol with the local investigators and their team. During the study, the completeness of the collected records will be checked based on the accuracy of entries, the adherence to the protocol and to Good Clinical Practice, the progress of data collection, and to ensure that source documents for each patient are properly stored. Validation procedures within the system will continuously check for data discrepancies, and the Principal Investigator at each site must certify that the data entered are complete and accurate. Data management will be done according to the internal procedures of clinical investigators and the organisation responsible for the study. Related information will be made available on request. All collected data will be reviewed for completeness, correctness and consistency. In case of issues, queries will be sent to the clinical site to complete, correct or comment on the data.

Data collection. Each clinical site will handle data in accordance with the applicable EU and national laws and the respective internal policies. Each clinical site will, thus, ensure, among others, the data's accuracy, completeness, and timeliness. Data that are derived from source documents must be consistent with the source documents, and discrepancies need to be justified in a documented rationale, signed and dated by the (principal) investigator and filed in the subject medical file. Any source documentation and any imaging sent to the organisation responsible for the study should have all subject identifiers removed and replaced with the subject's study ID.

Monitoring procedures. Monitoring visits (physical or remote) may be conducted before, during, and at the study's closure. The organisation responsible for the study for each site shall determine the frequency and timing of monitoring visits based on the scope of collected data, study compliance, and findings from previous visits.

The monitoring strategy covers the actions mentioned below (Table 7).

Table 7. Monitoring strategy

Actions	Parties involved	Methods to be used	Rationale for their use
Communication with stakeholders: 1) Clinical sites; 2) Bioethical Committee	Organisation responsible for the study and study team members, Bioethical Committee contact persons	Emails/ calls, visits (as appropriate to the specific issue(s) that trigger the communication with stakeholders).	Communication with stakeholders helps to ensure that the study is conducted as planned (in full scope and related time frames) and that all changes are well managed.
Monitoring visits: interim visits.	Organisation responsible for the study and	Onsite/remote monitoring visits could be conducted.	Interim Monitoring Visits may be conducted throughout the study to verify that:

Funded by the European Union

Not less than once per 3 months.	study team members		▪ The clinical site is conducting the study following applicable requirements, including the protocol, related procedures, and applicable regulatory requirements; ▪ Participant's safety, rights, and well-being are being protected; ▪ Recorded data are accurate, complete, and verifiable from source documentation.
Monitoring visits: For-cause visits (by request)	Organisation responsible for the study and study team members	These visits may involve either on-site monitoring or remote monitoring as appropriate to the specific issue(s) that trigger the visit.	For-cause visits will be conducted as applicable to address any unanticipated issues that arise in situations in which the site requires assistance. For-cause visits may be requested by the clinical site.
Monitoring visits: Close-out visit. Not later than 30 (thirty) working days after the clinical site approval that the study is implemented.	Organisation responsible for the study and study team members	The Close-Out Visit may be conducted either remotely or on-site.	A Close-Out Visit will be conducted to ensure that all study data and other study documentation are complete and accurate and that all study records have been reconciled.

*Monitoring visits could be performed remotely.

Study deviations and clinical study protocol changes. The clinical site is not allowed to deviate from the Clinical Study Protocol except with prior approval and under emergency circumstances. All deviations shall be documented and explained, regardless of the reason for the deviation. The clinical site shall obtain documented approval from the organisation responsible for the clinical study before implementation for any change in or deviation from the Clinical Study Protocol. In case of study deviations that can affect the subject's rights, safety, and well-being or the scientific integrity of the clinical study, approval from the Bioethical Committee/ other regulatory authority must also be obtained before implementation.

Study suspension or early termination. The study may be terminated or suspended at the initiative of the investigators if any of the following reasons arise:

- Data Privacy Concerns: If there are concerns regarding patient privacy and data protection, it may lead to the suspension or termination of the protocol. This could occur if there are breaches in data security, unauthorised access to patient records, or non-compliance with data protection regulations.
- Legal or Regulatory Issues: If there are legal or regulatory violations related to the study, such as non-compliance with institutional policies, local regulations, or applicable laws, the protocol procedures may be suspended or terminated to address these issues.
- External Factors: External circumstances such as natural disasters, public health emergencies, or unforeseen events that disrupt the healthcare system or impede data access and retrieval from EHRs may necessitate suspending or terminating the protocol procedures.

In this case, the clinical site must inform the Organisation responsible for the study of the reasons for terminating the study, and the data collected prior to terminating the study must be passed on to the organisation responsible for the study.
Any changes will be agreed upon with the Bioethical Committee that authorised the study.

Study close out. Organisation responsible for the study will notify the site of the intention to close the study. Study close-out visits may be performed. During these visits, the monitors will ensure that the clinical site's regulatory files are up to date-and complete and that any outstanding issues from previous visits have been resolved. Organisation responsible for the study will notify and inform the site(s) that all requirements have been met with a study closure letter.
The organisation responsible for the study will notify the Bioethical Committee about the closure by providing a Prospective Study report based on the Bioethical Committee/ other regulatory authority form.

Funded by the European Union

12. BIBLIOGRAPHY

[1] Y. Xi and P. Xu, "Global colorectal cancer burden in 2020 and projections to 2040," Transl. Oncol., vol. 14, no. 10, p. 101174, Oct. 2021, doi: 10.1016/j.tranon.2023.101174
[2] G. Argilés et al., "Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†," Ann. Oncol., vol. 31, no. 10, pp. 1291–1305, Oct. 2020, doi: 10.1016/j.annonc.2020.06.022.
[3] M. Riihimäki, H. Thomsen, K. Sundquist, J. Sundquist, and K. Hemminki, "Clinical landscape of cancer metastases," Cancer Med., vol. 7, no. 11, pp. 5534–5542, Oct. 2018, doi: 10.1002/cam4.1697.
[4] D. E. O'Sullivan et al., "Combinations of modifiable lifestyle behaviours in relation to colorectal cancer risk in Alberta's Tomorrow Project," Sci. Rep., vol. 10, no. 1, Art. no. 1, Nov. 2020, doi: 10.1038/s41598-020- 76294-w.
[5] V. Vymetalkova, K. Cervena, L. Bartu, and P. Vodicka, "Circulating Cell-Free DNA and Colorectal Cancer: A Systematic Review," Int. J. Mol. Sci., vol. 19, no. 11, p. 3356, Oct. 2018, doi: 10.3390/ijms19113356
[6] O. Mazouji, A. Ouhajjou, R. Incitti, and H. Mansour, "Updates on Clinical Use of Liquid Biopsy in Colorectal Cancer Screening, Diagnosis, Follow-Up, and Treatment Guidance," Front. Cell Dev. Biol., vol. 9, 2021, Accessed: May. 17, 2023. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcell.2021.660924
[7] A. Loktionov, "Biomarkers for detecting colorectal cancer non-invasively: DNA, RNA or proteins?," World J. Gastrointest. Oncol., vol. 12, no. 2, pp. 124–148, Feb. 2020, doi: 10.4251/wjgo.v12.i2.124.
[8] C. Pauli et al., "A Challenging Task: Identifying Patients with Cancer of Unknown Primary (CUP) According to ESMO Guidelines: The CUPISCO Trial Experience," The Oncologist, vol. 26, no. 5, pp. e769–e779, May 2021, doi: 10.1002/onco.13744.
[9] E. Rassy, T. Assi, and N. Pavlidis, "Exploring the biological hallmarks of cancer of unknown primary: where do we stand today?," Br. J. Cancer, vol. 122, no. 8, Art. no. 8, Apr. 2020, doi: 10.1038/s41416-019-0723-z.
[10] W. Zhao, M. Song, J. Zhang, M. Kuerban, and H. Wang, "Combined identification of long non-coding RNA CCAT1 and HOTAIR in serum as an effective screening for colorectal carcinoma," Int. J. Clin. Exp. Pathol., vol. 8, no. 11, pp. 14131–14140, Nov. 2015.
[11] R. Forghani, P. Savadjiev, A. Chatterjee, N. Muthukrishnan, C. Reinhold, and B. Forghani, "Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology," Comput. Struct. Biotechnol. J., vol. 17, pp. 995–1008, Jul. 2019, doi: 10.1016/j.csbj.2019.07.001.
[12] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
[13] P. Sabol et al., "Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images," J. Biomed. Inform., vol. 109, p. 103523, Sep. 2020, doi: 10.1016/j.jbi.2020.103523.
[14] F. Bénard, A. N. Barkun, M. Martel, and D. von Renteln, "Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations," World J. Gastroenterol., vol. 24, no. 1, pp. 124–138, Jan. 2018, doi: 10.3748/wjg.v24.i1.124.

13.LIST OF TABLES

14. LIST OF FIGURES

15. ANNEXES

ANNEX No. 1. A INFORMED CONSENT FORM

The consent form below provides for the lawful participation of individuals in the prospective study and for the lawful processing of the respective personal data. To this end, the present consent form is largely based on the requirements set out under the General Data Protection Regulation (GDPR), as well as under other EU laws, such as the Medical Devices Regulation (MDR) and the ethical principles for medical research involving human subjects set forth under the Helsinki Declaration.

INFORMED CONSENT FORM

Title of study: ‘’Prospective study data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms’’
Protocol No.: 1
Organisation responsible for the study:
Address: Tel.: Email:
Representative of the organisation responsible for the prospective study: Local representative:
Clinical site:
Address: Tel.: Email:
Participant ID:

PURPOSE OF THIS DOCUMENT

By signing this document, you agree to participate in the prospective study. Read this document carefully, if you do not understand any word or statement, be sure to ask the researcher/other person

authorised by the Organisation responsible for the study any questions you may have. You can talk to family, friends, or your doctor before making a decision.

GENERAL INTRODUCTION

If you have been invited to take part in this study, it is because you are about to undergo a colonoscopy as part of a colorectal cancer screening program.

1. Colorectal cancer

As the name suggests, colorectal cancer develops in the colon, also known as the large intestine, or in its last part, the rectum. The starting point of colorectal cancer is a protruding growth of tissue from the intestinal wall called a polyp. Although in most cases, polyps are non-cancerous (benign), some are precancerous lesions that can give rise to a tumour. The evolution of precancerous polyps into tumours can take 10 to 15 years as they go through several slow stages of transformation.

Colorectal cancer is the third most common cancer in men and the second in women, accounting for 10% of all cancers worldwide. It ranks second in terms of cancer-related mortality, just behind lung cancer.

2. Colorectal cancer screening

In [region or country*], a colorectal cancer screening program has been in operation since [year or exact time] for people aged between 50 and 74. This screening can be carried out at home, using a test based on the search for occult blood in a stool sample. Although this method is simple to perform, it only detects the presence or absence of blood in the stool — it can't determine what's causing the bleeding. If blood is detected through a faecal occult blood test, additional tests may be needed to determine the source of the bleeding. Therefore, although more invasive, colonoscopy remains the most reliable method of screening for colorectal cancer in patients with a positive faecal occult blood test, as it enables polyps and other lesions to be visualised and removed using an endoscope equipped with a camera. The risk of colorectal cancer following colonoscopy has been shown to be reduced by 70-90%. Early detection and removal of a pre-cancerous polyp prevents its progression to cancer. In this way, colonoscopy saves many lives.

Nevertheless, although colorectal cancer is now considered an easily preventable disease thanks to screening, long waiting and preparation times for colonoscopy prevent the implementation of large-scale screening for systematic surveillance and follow-up.

Funded by the European Union

Since the 1990s, there has been a gradual increase in the rate of colorectal cancer in adults under the age of 50. Although the reasons for this are still largely unknown, it has been suggested that environmental and behavioural changes influencing the microbiome, along with familial predisposition, are at the root of colorectal cancer in people under 50. Colonoscopy reimbursement in your country if applicable.

Therefore, the need to develop a large-scale, inexpensive, and non-invasive method of early detection of colorectal cancer is urgent.

*All the highlighted parts should be adjusted according to the individual circumstances of the clinical partners.

MAIN OBJECTIVE OF THE STUDY

You are invited to participate in part of a project called DIOPTRA. DIOPTRA - Early Dynamic Screening For Colorectal Cancer Via Novel Protein Biomarkers Reflecting Biological Initiation Mechanisms – is a 4-year project co-funded by the European Union's Horizon Europe Research and Innovation Programme under the grant agreement No 101096649, Swiss State Secretariat for Education, Research and Innovation and the UK Research and Innovation which aims to develop a routine blood test accessible to all ages, to identify people who would not otherwise be screened according to current European or national guidelines.

The previous part of the DIOPTRA project would have identified in around 200 participants a protein group whose quantity varies during a precancerous stage of colon cancer. By quantifying this group of proteins, this blood test will identify citizens who should undergo further colonoscopy screening. To validate this method, you can give a blood sample during your colonoscopy visit at the Hospital's gastroenterology department. Once validated, this blood test has many advantages: it is almost non-invasive, inexpensive and could be well accepted by most of the population. As a result, DIOPTRA is positioning itself in the increasingly personalised medicine of the future, capable of adapting to the particularities of each individual.

OTHER OBJECTIVES OF THE STUDY

In addition to an early detection method for colon cancer, numerous scientific studies have identified parameters called "risk factors" that could be associated with the development of colorectal cancer. Their importance is not negligible. These risk factors can generate suggestions for daily habits and may be very useful in the prevention of colorectal cancer.

Another aim of the study is to validate some of the risk factors identified in the previous part of the project. To do this, we'll need to collect some information about you: socio-demographic data (age, sex, height, weight, occupation, education, standard of living, etc.), data and results of medical examinations (e.g., family history, colonoscopy diagnosis, medication, blood test results, etc.), behavioural information (cigarettes, alcohol), nutritional habits, physical activity, etc.

The findings and the final DIOPTRA solution will be the subject of a study of healthcare performance indicators, with the goal of widening screening eligibility thanks to an effective, minimally invasive, and financially affordable method.

STUDY PROCEDURES

The study will be conducted only on the day of your colonoscopy visit:

Before inviting you to participate in this study, the healthcare professionals at the hospital will consult your medical file to ensure that you meet all the criteria for participation.

During this initial phase, your participation will be limited to:

A. A donation of two blood samples of 10 ml each

B. On the day of your colonoscopy, answer questions about your socio-demographic, behavioural information, habits, physical activities, etc, which will be asked by a health professional [adjust to local method].

C. Sharing your data with the other partners of the DIOPTRA project in a pseudonymised format for the study as mentioned above.

You will then undergo your colonoscopy as planned.

This study will include at least 1,600 participants from countries participating in DIOPTRA, of whom at least 200 will be enrolled at [Hospital].

To participate in this study, the medical staff will make sure that you meet the following criteria:

- Showing any indication for total colonoscopy (including routine screening and presence of symptoms/ FIT positive).
- Age between 18-80 years at the moment of recruitment (see above).
- Absence of significant comorbidities (ASA IV).

- Ability to provide valid (written informed) consent.
- Boston-Bowel-Preparation-Scale (BBPS) left/transverse/right colon score ≥2, total score ≥6

RISKS AND INCONVENIENCES

You will not experience any inconvenience by participating in this study. This study will not impact the treatment you have been offered or the diagnostic and monitoring procedures of the usual medical practice in your clinical case. Blood sampling may (rarely) cause pain, bleeding, bruising, or localised infection at the blood sampling site. In addition, some people may feel dizzy or faint during the procedure.

You must be aware that any study or blood sampling may involve certain risks, as with the standard treatment you receive. However, the researcher and all the study team members will do everything possible and necessary to ensure these risks are minimal.

Given that blood samples will be collected using a minimally invasive method and the study does not pose any additional risks, it is not covered by the civil liability insurance of the clinical site and the organisation responsible for the study.

BENEFITS

You will not derive any direct benefit (medical, financial, or otherwise) from participating in this study. The donation of human body material samples is free of charge, and there will be no financial compensation if the research results in new medical treatments.

However, in general, early detection of CRC can significantly enhance survival rates and treatment outcomes. Medical specialists could offer personalised prevention plans to reduce CRC risk.

TERMINATION OF PARTICIPATION IN THIS STUDY

You participate in the study voluntarily, so you have the right to opt-out, and once you start, you can withdraw from it at any time without giving reasons and without any resulting detriment. If you cannot decide on further access to the study due to your deteriorating health, this decision can be taken by your spouse or, if not, one of your parents, adult children, or another legal representative.

Funded by the European Union

Your decision not to participate or to terminate your participation in the study will not affect the routine health care provided.

Your participation in the study will be automatically terminated if you no longer belong to the subject inclusion criteria, do not come to the scheduled visits, or do not follow the investigators' instructions.

WILL YOU INCUR ANY COSTS IN PARTICIPATING IN THIS STUDY?

You will not incur any cost for participating in the study as your visits will be part of the routine healthcare service the responsible organisation has offered you. You will not be charged for any visits, consultations, examinations, or treatments specific to this study. Normal medical expenses (unrelated to the study), even if generated on the same day, will be billed to you (and/or your insurance company) as usual.

MANAGEMENT OF PERSONAL DATA

As part of your participation in the study, your data will also be processed. We ask your consent to collect, process, and store your personal data and your body material for the purpose of this study, as described in detail in this section. This section also explains how you can exercise your data protection rights in accordance with the EU General Data Protection Regulation 2016/679 ('GDPR'), [*the national data protection law applicable to the organisation*] and other related laws and regulations.

Who is your Data Controller?

[*Organisation responsible for the study*] is the data controller for your personal data that will be processed for the study. The contact details can be found on the first page of this consent form.

The Data Controller has its own data protection officer (‘DPO’) who oversees compliance with the applicable data protection and privacy laws and functions as a point of contact for all privacy-related queries. If you have any queries regarding the protection of your personal data, you may directly contact DPO at [*email address or phone number of DPO*].

What personal data and body material do we process?

In this study, we collect, process, and store the following categories and types of your personal data: your name, gender, weight, height, date of birth, medical file number and information available in your medical file kept at the clinical site, including but not limited to colonoscopy results, genetic data, clinical diagnosis, prescribed medications, allergies and self-reported symptoms, your answers to the

questionnaire related to demographic, dietary, financial, lifestyle and behavioural data and your habits corresponding to potential risk factors for colorectal cancer and medical information including personal and family history along with symptoms that we collect during the study via the "Medical Information History Form". As for body material, we only collect, use and store a 20mL blood sample in two analysis tubes.

Who can access your personal data and body material?

*Your personal data will be accessed and processed by the study team of [*organisation responsible for the study*].* After the pseudonymisation of your personal data, the study team may share your data in a pseudonymous (coded) form with other organisations participating in the DIOPTRA project for the purpose of the study in accordance with the DIOPTRA Framework data processing agreement dated 26 August 2024. Your body material will be used by the study team in charge of managing your bold material in order to measure the concentration of a panel of protein biomarkers in your blood. The study team who uses your samples may also receive the personal data linked to the samples they need for their research. The study team has a duty of confidentiality with regard to the body materials and the personal data collected.

To verify the quality of the study or for regulatory compliance purposes, your data may be examined by third parties (for example, competent national and European authorities, including ethics committees, health authorities, and external auditors). In any event, this may only be done under the supervision of the study team in charge within [*organisation responsible for the study*] or the physician managing the human body material at the biobank and/or by any other authorised persons bound by the obligation of professional secrecy and confidentiality.

Will your personal data be transferred to countries outside the European Union/European Economic Area?

Your data may be transferred to Cambridge Medical Academy LTD, established in Cambridge, United Kingdom, at Bay 13 Clifford Albutt Building Hills Road, with VAT number GB249759350, which is a partner of the DIOPTRA consortium. The transfer of your data to the United Kingdom will be done according to the European Commission Implementing Decision (EU) 2021/1772) on the adequate protection of personal data by the United Kingdom and the DIOPTRA Framework data processing agreement, dated 26 August 2024, executed between Cambridge Medical Academy LTD and [organisation responsible for the study].

How do we protect your privacy?

To protect your personal data, your identity information will be replaced by a code, and together with your body material, your personal data will be stored in a pseudonymous (coded) form. We keep the

key to the code in a safe place on the clinical site. When we process your data and body material, we always use only that code. In addition, all necessary measures are taken to protect the confidentiality and security of your encoded data in accordance with the applicable legislation at the European and national levels. Finally, we have executed a data processing agreement with the recipient organisations of your data to ensure that appropriate technical and organisational measures are implemented to protect your data throughout each recipient's processing operations.

For how long do we store your personal data and body material?

Your personal data and body materials will be kept for the duration of the DIOPTRA project (maximum four (4) years starting from the time of the data collection), or the time may be required by other applicable laws to this study, whichever comes later. If they are no longer needed for the purpose of the study or for compliance with other applicable laws, we will erase your personal data and destroy body samples before this date.

Whether we do automated decision-making or profiling

The study deploys DIOPTRA advanced Artificial intelligence-based cancer screening system, which is developed by the DIOPTRA project to analyse your personal data and information obtained from your body materials to carry out colorectal cancer risk assessment, screening, and progression based on patients' profiles in general. However, any automated decision, including profiling generated by DIOPTRA's advanced Artificial Intelligence based cancer screening system, will not affect you nor have any impact on the healthcare service you receive from *[organisation responsible for the study]*.

What happens if there are coincidental findings?

It is possible that during the study, we discover something that is not directly relevant to the study but is important to your health or to the health of your family members. In that case, the study team, including your physician, will be informed. Under no circumstances can any coincidental findings be considered as results that can be used to make a medical diagnosis. The study team will, therefore, decide whether it is useful to communicate this information to you and whether to offer you, for instance, advice, requests, complementary examinations, or treatments totally independent of the present study. This information may be of benefit to you in terms of your health, but in some cases, it may also cause you anxiety or other psychological difficulties.

What data protection rights do you have?

You have the right to have access to all study information concerning you and to request, if necessary, rectification, data portability and to restrict processing of your personal data. You have a right to withdraw your consent for the use of your personal data at any time. Please inform the study team if you

wish to do so. Note that if you withdraw your consent, we will cease processing your personal data immediately. The study team will destroy your body material and erase your personal data after you withdraw your consent. If, however, assessments with your body material have been carried out prior to the withdrawal of your consent, the study team may continue to use the results from such assessments, provided that such assessments do not contain your personal data. Do you want to know more about your rights when processing personal data? Visit [website].

CONTACT TO AUTHORITIES

For your rights as a study participant, you can apply to the Ethics Committee, which has given us a permit to conduct this study. To exercise your data protection rights, please directly contact [*the Data Protection Officer of the organisation responsible for the study*] [contact details, e.g., email, phone, etc., and website].

If you have any complaints about the processing of your personal data, we recommend that you first discuss them with the study team or directly contact the [*Data Protection Officer of the organisation responsible for the study*]. You can also submit a complaint to the national Data Protection Authority [Insert the full name and contact details of the national data protection authority in the country where the study is carried out].

CONSENT TO PARTICIPATE IN THE PROSPECTIVE STUDY

By signing this information and consent form, I hereby certify that:

- I have read this Informed Consent Form and have understood the information about the nature, objectives, benefits, implications, risks, and inconveniences of the study, the use of my body materials, its purpose, how it is carried out, and what is expected of me. I was given the opportunity to ask questions and received satisfactory answers. I have had enough time to decide if I want to take part calmly.

- I have filled in this informed consent form of my own free will and without being subject to any inappropriate pressure or influence by the researcher or by a member of the study team.

- I understand that participation in the study is voluntary. I also know that at any time I can withdraw from the study at any time without giving any reason [1].

- I understand that in order to withdraw my consent to participate in the study, I must inform the researcher / other person authorised by the clinical site identified below in writing.
- I renounce any rights whatsoever over my body materials collected within the study and the results of the study to be carried out with these samples to the extent permitted by applicable law.
- I have been informed about the processing of my personal data for the purpose of this study, including types of personal data to be processed, the data controller, potential recipients of my personal data, data security measures, and my data protection rights, including my right to withdraw my consent to the processing at any time.
- I confirm that I have received a copy of the Informed Person Consent Form, signed by the researcher / other clinical site authorised person

To give your free and explicit consent, please tick yes or no in the table below:

I give my consent to participate in this study	Yes ☐	No☐
I give my explicit consent to the processing of my personal data, including special categories of personal data, for the purpose of this study, as stated herein.	Yes ☐	No☐
I agree to receive, via my referring physician, the information generated by the study or research on my body material samples of significant importance or potential interest for my state of health.	Yes ☐	No☐

Person (or other person with the right to give consent on behalf of the participant)										
								MMMM-mm-dd		_:_

name		surname		Representation basis		signature		Signing date		Signing time

I confirm that I have provided information about the study to the person mentioned above.

I confirm that the person (or other person entitled to give consent) has been given sufficient time to decide to participate in the study, taking into account the nature of the clinical study, as well as considering other circumstances that may influence the decision.

I encouraged the person (or other person with the right to consent) to ask questions and answer them.

Researcher / other person authorised by the clinical site.										
								MMMM-mm-dd		_:_
name		surname		duties in the study		signature		Signing date		Signing time

[1] If the consent to participate in the study is given by the person himself

ANNEX No. 1. B INFORMED CONSENT FORM

The consent form below provides for the lawful participation of individuals in the prospective follow-up study and for the lawful processing of the respective personal data. To this end, the present consent form is largely based on the requirements set out under the General Data Protection Regulation (GDPR), as well as under other EU laws, such as the Medical Devices Regulation (MDR) and the ethical principles for medical research involving human subjects set forth under the Helsinki Declaration.

INFORMED CONSENT FORM

Title of study: ‘’Prospective follow-up study data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms’’
Protocol No.: 1
Organisation responsible for the study:
Address: Tel.: Email:
Representative of the organisation responsible for the prospective study: Local representative:
Clinical site:
Address: Tel.: Email:
Participant ID:

PURPOSE OF THIS DOCUMENT

By signing this document, you agree to participate in the prospective follow-up study. Read this document carefully. If you do not understand any word or statement, be sure to ask the researcher/other

person authorised by the Organisation responsible for the study any questions you may have. You can talk to family, friends, or your doctor before making a decision.

GENERAL INTRODUCTION

If you have been invited to take part in this study, it is because you are about to undergo a colonoscopy as part of a colorectal cancer screening program.

1. Colorectal cancer

As the name suggests, colorectal cancer develops in the colon, also known as the large intestine, or in its last part, the rectum. The starting point of colorectal cancer is a protruding growth of tissue from the intestinal wall called a polyp. Although in the majority of cases, polyps are non-cancerous (benign), some are precancerous lesions and can give rise to a tumour. The evolution of precancerous polyps into tumours can take 10 to 15 years as they go through several slow stages of transformation.

Colorectal cancer is the third most common cancer in men and the second in women, accounting for 10% of all cancers worldwide. It ranks second in terms of cancer-related mortality, just behind lung cancer.

2. Colorectal cancer screening

In [region or country], a colorectal cancer screening program has been in operation since [year or exact time] for people aged between 50 and 74. This screening can be carried out at home, using a test based on the search for occult blood in a stool sample. Although this method is simple to perform, it only detects the presence or absence of blood in the stool — it can't determine what's causing the bleeding. If blood is detected through a faecal occult blood test, additional tests may be needed to determine the source of the bleeding. Therefore, although more invasive, colonoscopy remains the most reliable method of screening for colorectal cancer in patients with a positive faecal occult blood test, as it enables polyps and other lesions to be visualised and removed using an endoscope equipped with a camera. The risk of colorectal cancer following colonoscopy has been shown to be reduced by 70-90%. Early detection and removal of a pre-cancerous polyp prevents its progression to cancer. In this way, colonoscopy saves many lives.

Nevertheless, although colorectal cancer is now considered an easily preventable disease thanks to screening, long waiting and preparation times for colonoscopy prevent the implementation of large-scale screening for systematic surveillance and follow-up.

Since the 1990s, there has been a gradual increase in the rate of colorectal cancer in adults under the age of 50. Although the reasons for this are still unknown, it has been suggested that environmental and behavioural changes influencing the microbiome are at the root of colorectal cancer in people under 50. Colonoscopy reimbursement in your country if applicable.

The need to develop a large-scale, inexpensive, and non-invasive method of early detection of colorectal cancer is therefore urgent.

MAIN OBJECTIVE OF THE STUDY

You are invited to participate in part of a project called DIOPTRA. DIOPTRA - Early Dynamic Screening For Colorectal Cancer Via Novel Protein Biomarkers Reflecting Biological Initiation Mechanisms – is a 4-year project co-funded by the European Union's Horizon Europe Research and Innovation Programme under the grant agreement No 101096649, Swiss State Secretariat for Education, Research and Innovation and the UK Research and Innovation which aims to develop a routine blood test accessible to all ages, in order to identify people who would not otherwise be screened according to current European or national guidelines.

The previous part of the DIOPTRA project would have identified in around 200 participants a protein group whose quantity varies during a precancerous stage of colon cancer. By quantifying this group of proteins, this blood test will be able to identify those citizens who absolutely should undergo further screening by colonoscopy. To validate this method, you are invited to give a blood sample during your colonoscopy visit at the Hospital's gastroenterology department. Once validated, this blood test has many advantages: it is almost non-invasive, inexpensive and could be well accepted by most of the population. As a result, DIOPTRA is positioning itself in the increasingly personalised medicine of the future, capable of adapting to the particularities of each individual.

OTHER OBJECTIVES OF THE STUDY

In addition to an early detection method for colon cancer, numerous scientific studies have identified parameters called "risk factors" which could be associated with the development of colorectal cancer, and their importance is not negligible. Suggestions for daily habits can be generated from these risk factors and may be very useful in the prevention of colorectal cancer.

Another aim of the study is to validate some of the risk factors identified in the previous part of the project. To do this, we'll need to collect some information about you: socio-demographic data (age, sex,

height, weight, occupation, education, standard of living, etc.), data and results of medical examinations (e.g. family history, colonoscopy diagnosis, medication, blood test results, etc.), behavioural information (cigarettes, alcohol) and nutritional habits, physical activity, etc. The DIOPTRA application would be created to help collect certain information, offer up-to-date, personalised suggestions and raise awareness of the early detection of colorectal cancer.

The findings and the final DIOPTRA solution will be the subject of a study of healthcare performance indicators in view of widening screening eligibility thanks to an effective, minimally invasive and financially affordable method.

STUDY PROCEDURES

The study will be conducted in two phases:

Phase I (day of the colonoscopy):

Before inviting you to take part in this study, the healthcare professionals at Hospital will have consulted your medical file to ensure that you meet all the criteria for participation in this study.

During this initial phase, your participation will be limited to:

A. A donation of two blood samples of 10 ml each

B. On the day of your colonoscopy, answer questions about your eating and exercise habits, which will be asked by a health professional[adjust to local method].

C. Sharing of your data with the other partners of the DIOPTRA project in a pseudonymised format for the study as mentioned above.

You will then undergo your colonoscopy as planned.

This first part of the study will include at least 1,600 participants from countries participating in DIOPTRA, of whom at least 200 will be enrolled at [Hospital].

In order to participate in the first part of the study, the medical staff will make sure that you meet the following criteria:

- Showing any indication for total colonoscopy (including routine screening and presence of symptoms/ FIT positive).

- Age between 18-80 years at the moment of recruitment (see above).
- Absence of significant comorbidities (ASA IV).
- Ability to provide valid (written informed) consent.
- Boston-Bowel-Preparation-Scale (BBPS) left/transverse/right colon score ≥2, total score ≥6.

Phase II:

You are invited to participate in this phase if you meet the following criteria:

- Meeting all the Phase I inclusion criteria above.
- Willing to use the DIOPTRA mobile application regularly.
- Showing an appropriate level of digital literacy allows the management of mobile terminals (smartphones, smartphone apps, tablets).
- Having a good coverage of internet connection at home.
- Having a smartphone/ tablet available for personal use (in order to be able to use the app).
- Belonging to the healthy or non-advanced adenoma groups based on the result of the corresponding colonoscopy visit.

This second phase of the study will involve 416 participants, including 52 from Hospital. If you met all Phase I criteria but were not eligible for Phase II, you could still agree to participate in Phase I only.

After your hospital visit in Phase I, you will be contacted at a later date to confirm your participation in the second phase of the study. You will be provided with a mobile application called DIOPTRA that will contain information on:

· colorectal cancer occurrence and symptoms,

· local colorectal cancer screening guidelines,

· factors that may affect the risk of an individual developing colorectal cancer,

· lifestyle suggestions that are known to decrease the risk of developing colorectal cancer,

· DIOPTRA project

Its easy-to-use interface will help you in:

- recognising potential symptoms for you or your family members,
- learning current recommended screening guidelines,
- maintaining a healthy lifestyle that potentially reduces the risk of colorectal cancer.

Moreover, you will be able to use this application to get healthy lifestyle suggestions that reduce the risk of colorectal cancer that are tailored to your own needs and health status. To accomplish this, you will only need to fill out a questionnaire within the application, which in turn will provide you with a personalised suggestion. On several occasions during the 12 months, you may be contacted by the application for some updated questions. You will also be offered the opportunity to be re-assessed after 12 months and check if you accept the suggestion and if a healthier lifestyle change has affected the risk of colorectal cancer. Therefore, you will be able to receive expert information on how your health has progressed, which will also be useful for you in the future.

Apart from using the application, you will be asked to return to the Hospital for a second visit 12 months after your colonoscopy. During this second visit, you will be asked to donate two blood samples of 10 ml each that will be later used to validate developed blood test for early onset colorectal cancer detection.

Your participation in Phase II of the study ends once you complete the blood donation and answer all questionnaires in the application 12 months after your first visit.

RISKS AND INCONVENIENCES

You will not experience any inconvenience by participating in this study. This study will not have any impact on the treatment you have been offered or the diagnostic and monitoring procedures of the usual medical practice in your clinical case. Blood sampling may (rarely) cause pain, bleeding, bruising, or localised infection at the blood sampling site. In addition, some people may feel dizzy or faint during the procedure.

You must be aware that any study or blood sampling may involve certain risks, the same as with the standard treatment you receive. However, the researcher and all the members of the study team will do everything possible and necessary to ensure that these risks are kept to a minimum.

Given that the collection of blood samples will be done with a minimally invasive method and the study does not pose any additional risks, the study is not covered by the civil liability insurance of the clinical site and the organisation responsible for the study.

Funded by the European Union

BENEFITS

You will not derive any direct benefit (medical, financial, or otherwise) from your participation in this study. The donation of samples of human body material is free of charge, and there will be no financial compensation if the research results in new medical treatments.

However, in general, early detection of CRC can significantly enhance survival rates and treatment outcomes. Medical specialists could offer personalised prevention plans to reduce CRC risk.

TERMINATION OF PARTICIPATION IN THIS STUDY

You participate in the study voluntarily, so you have the right to opt-out, and once you start, you can withdraw from it at any time without giving reasons and without any resulting detriment. If you are unable to decide on further access to the study due to your deteriorating health, this decision will be taken by your spouse or, if not, one of your parents, adult children, or another legal representative.

Your decision not to participate or to terminate your participation in the study will not affect the routine health care provided.

Your participation in the study will be automatically terminated if you no longer belong to the subject inclusion criteria if you do not come to the scheduled visits, or if you do not follow the investigators' instructions.

WILL YOU INCUR ANY COSTS IN PARTICIPATING IN THIS STUDY?

You will not incur any cost for participating in the study as your visits will be part of the routine healthcare service you have been offered by the responsible organisation. You will not be charged for any visits, consultations, examinations, or treatments specific to this study. Normal medical expenses (not related to the study), even if generated on the same day, will be billed to you (and/or your insurance company) as usual.

MANAGEMENT OF PERSONAL DATA

As part of your participation in the study, your personal data will also be processed. We ask your consent to collect, process, and store your personal data and your body material for the purpose of this study, as

described in detail in this section. This section also explains how you can exercise your data protection rights in accordance with the EU General Data Protection Regulation 2016/679 ('GDPR'), [*the national data protection law applicable to the organisation responsible for the clinical study*] and other related laws and regulations.

Who is your Data Controller?

[*organisation responsible for the study*] is the Data Controller of your personal data that will be processed for the clinical study. The contact details can be found on the first page of this consent form.

The Data Controller has its own data protection officer ('DPO') who oversees compliance with the applicable data protection and privacy laws and functions as a point of contact for all privacy-related queries. If you have any queries regarding the protection of your personal data, you may directly contact DPO at [*email address or phone number of DPO*].

What personal data and body material do we process?

In this project, we collect, process, and store the following categories and types of your personal data: your name, gender, weight, height, date of birth, medical file number and information available in your medical file kept at the clinical site, including but not limited to colonoscopy results, genetic data, clinical diagnosis, prescribed medications, allergies, self-reported symptoms and other data related to your health, medical information including personal and family history along with symptoms that we collect during the study via the "Medical Information History Form".

The mobile app is used to process the following categories and types of personal data:

CATEGORY	**DETAILS**
Identification and Registration	User ID number.
Device and Connection Information	IP number, mobile identification number.
Sociodemographic	Age, Gender, Country of Birth, Place of Residency, Weight, Height, level of education, income, and employment details.
Lifestyle	Smoking habits, alcohol consumption, and physical activity level.

Dietary Information	Consumption of fruits and vegetables, processed meat, dairy products, white meat, whole grains, sugary drinks and desserts, fast food, intake of dietary supplements including omega 3, multivitamins, vitamins B6, C, D, magnesium, calcium, iron, probiotics, fibre supplements, folic acid.
Stress Level	Perceived stress, including your feelings and thoughts.

As for body material, we only collect, use and store a 20mL blood sample in two analysis tubes in the first phase. If you also participate in the second phase of the study, we will also collect and analyse an additional 20mL blood sample.

Who can access your personal data and body material?

Your personal data will be accessed and processed by the study team of *[*organisation responsible for the study*].* We may share your personal data in a pseudonymous (coded) form with other organisations participating in the DIOPTRA project and National Infrastructures for Research and Technology (GRNET) S.A. as a sub-processor for further processing of your data for the purpose of the DIOPTRA research in accordance with the DIOPTRA Framework data processing agreement, dated 26 August 2024.

Your body material will be used by the study team in charge in order to measure the concentration of a panel of protein biomarkers in your blood. The study team who uses your samples may also receive the personal data linked to the samples they need for their research. The study team has a duty of confidentiality with regard to the body materials and the personal data collected.

To verify the quality of the study or for regulatory compliance purposes, your data may be examined by third parties (for example, competent national and European authorities, including ethics committees, health authorities, and external auditors). In any event, this may only be done under the supervision of the study team in charge within [*organisation responsible for the study*] or the physician managing the human body material at the biobank and/or by any other authorised persons bound by the obligation of professional secrecy and confidentiality.

Will your personal data be transferred to countries outside the European Union/European Economic Area?

Your data may be transferred to Cambridge Medical Academy LTD, established in Cambridge, United Kingdom, at Bay 13 Clifford Albutt Building Hills Road, with VAT number GB249759350, which is a partner of the DIOPTRA consortium. The transfer of your data to the United Kingdom will be done according to the European Commission Implementing Decision (EU) 2021/1772) on the adequate protection of personal data by the United Kingdom and the DIOPTRA Framework data processing

agreement, dated 26 August 2024, executed between Cambridge Medical Academy LTD and [organisation responsible for the study].

How do we protect your privacy?

To protect your personal data, your identity information in your medical file will be replaced by a code. The Data Controller will assign you a unique User ID for the mobile app, and you will provide your data in the mobile app under your User ID. Together with your body material, your personal data will be stored in a pseudonymous (coded) form. We keep the key to the code and the User ID in a safe place on the clinical site. In addition, all necessary measures are taken to protect the confidentiality and security of your encoded data in accordance with the applicable legislation at the European and national levels. We also apply authentication-based access control to the datasets and servers containing your personal data and implement the need-to-know principle in the vetting of any researcher involved in the processing of personal data. Only authorised persons will have access to your personal data in order to monitor and carry out the processing activities.

Finally, we have executed a data processing agreement dated 26 August 2024 with the recipient organisations of your data to ensure that appropriate technical and organisational measures are implemented to protect your data throughout each recipient's processing operations.

For how long do we store your personal data and body material?

We aim to pseudonymise your personal data upon collection and store it in the local servers of the Data Controller as well as in the hosting infrastructure of GRNET. Your body materials will be stored in [*please indicate where body materials will be stored*]. Your personal data and body materials will be kept for the duration of the DIOPTRA project (maximum four (4) years starting from the time of the data collection), or the time may be required by other applicable laws to this study, whichever comes later. If they are no longer needed for the purpose of the study or for compliance with other applicable laws, we will erase your personal data and destroy body samples before this date.

Whether we do automated decision-making or profiling

The DIOPTRA study deploys Artificial Intelligence systems to process your personal data in order to create your risk profiles and wellness score. Your risk profile and wellness score will be shared with your clinicians to provide personalised health, wellness and lifestyle recommendations. However, any automated decision, including recommendations and profiles, will not have a significant impact on the healthcare service you receive from *[organisation responsible for the study]*.

What happens if there are coincidental findings?

It is possible that during the study, we discover something that is not directly relevant to the study but is important to your health or to the health of your family members. In that case, the study team, including your physician, will be informed. Under no circumstances can any coincidental findings be considered as results that can be used to make a medical diagnosis. The physician will, therefore, decide whether it is useful to communicate this information to you and whether to offer you, for instance, advice, request, complementary examinations, or treatments totally independent of the present study. This information may be of benefit to you in terms of your health, but in some cases, it may also cause you anxiety or other psychological difficulties.

What data protection rights do you have?

You have the right to have access to all study information concerning you and to request, if necessary, rectification, data portability and to restrict processing of your personal data. You have a right to withdraw your consent for the use of your personal data at any time. Please inform the study team if you wish to do so. Note that if you withdraw your consent, we will cease processing your personal data immediately. The study team will destroy your body material and erase your personal data after you withdraw your consent. If, however, assessments with your body material have been carried out prior to the withdrawal of your consent, the study team may continue to use the results from such assessments, provided that such assessments do not contain your personal data. Do you want to know more about your rights when processing personal data? Visit [website].

CONTACT TO AUTHORITIES

For your rights as a study participant, you can apply to the Ethics Committee, which has given you a permit to conduct this study. To exercise your data protection rights, please directly contact [*Data Protection Officer of the organisation responsible for the study*] [contact details, e.g., email, phone, etc., and website].

If you have any complaints about the processing of your personal data, we recommend that you first discuss them with the study team or directly contact the [*Data Protection Officer of the organisation responsible for the study*]. You can also submit a complaint to the national Data Protection Authority [Insert the full name and contact details of the national data protection authority in the country where the study is carried out].

CONSENT TO PARTICIPATE IN THE PROSPECTIVE STUDY

By signing this information and consent form, I hereby certify that:

- I have read this Informed Consent Form and have understood the information about the nature, objectives, benefits, implications, risks, and inconveniences of the study, the use of my body materials, its purpose, how it is carried out, and what is expected of me. I was given the opportunity to ask questions and received satisfactory answers. I have had enough time to decide if I wanted to take part calmly.

- I have filled in this informed consent form of my own free will and without being subject to any inappropriate pressure or influence by the researcher or by a member of the study team.

- I understand that participation in the study and use of the mobile app is voluntary. I also know that at any time I can withdraw from the study and stop using the mobile app at any time without giving any reason [1].

- I understand that in order to withdraw my consent to participate in the study, I must inform the researcher / other person authorised by the clinical site identified below in writing.

- I renounce any rights whatsoever over my body materials collected within the study and the results of the study to be carried out with these samples to the extent permitted by applicable law.

- I have been informed about the processing of my personal data for the purpose of this study, including types of personal data to be processed, the data controller, potential recipients of my personal data, data security measures, and my data protection rights, including my right to withdraw my consent to the processing at any time.

- I confirm that I have received a copy of the Informed Person Consent Form, signed by the researcher / other clinical site authorised person.

To give your free and explicit consent, please tick yes or no in the table below:

I give my consent to participate in this study (If I am not eligible for participation in Phase II of the study, my consent will only be valid for Phase I)	Yes □	No□

I give my explicit consent to the processing of my personal data, including special categories of personal data, for the purpose of this study, as stated herein.	Yes ☐	No☐
I agree to be contacted by the study team to participate in the second phase of the study after my first colonoscopy visit.	Yes ☐	No☐
I consent to my personal data being processed with artificial intelligence in order to create my colorectal cancer risk profile and wellness score, which will be shared with my clinicians and to provide me with recommendations	Yes ☐	No☐
I agree to receive, via my referring physician, the information generated by the study or research on my body material samples of significant importance or potential interest for my state of health.	Yes ☐	No☐

Person (or other person with the right to give consent)										
								MMMM-mm-dd		_:_
name		surname		Representation basis		signature		Signing date		Signing time

I confirm that I have provided information about the study to the person mentioned above.

I confirm that the person (or other person entitled to give consent) has been given sufficient time to decide to participate in the study, taking into account the nature of the clinical study, as well as considering other circumstances that may influence the decision.

I encouraged the person (or other person with the right to consent) to ask questions and answered them.

Funded by the European Union

Researcher / other person authorised by the clinical site.										
								MMMM-mm-dd		_:_
name		surname		duties in the study		signature		Signing date		Signing time

[1] If the consent to participate in the study is given by the person himself

ANNEX No. 2. BEHAVIOURAL QUESTIONNAIRE

Sociodemographic (mark with an x)	
What is your age? (Write a number in years)	
Sex	☐ Male ☐ Female ☐ Not wish to answer
What is your country of birth?	
What is the country where you took the most of your life?	
Weight (in kg)	
Height (in cm)	
Where have you usually resided for most of your life?	☐ Urban[1] ☐ Rural[2]
What is the highest level of education you have obtained?	☐ Primary school graduate ☐ Middle school graduate ☐ High school graduate ☐ College or University degree ☐ Post-graduate degree ☐ None of them
What is your monthly net income? Which of the following best describes how you feel about your household income today?	☐ Living comfortably on current income ☐ Current income is something to live with ☐It is difficult to live on the current income
What is your work situation?	☐ Employed ☐ Unemployed ☐ Student/not currently employed ☐ Retired
Which sector your main work belong to?[3]	☐ Healthcare/Medical ☐ Education ☐ Science and Research ☐ Public Services/Government ☐ Arts/Creative ☐ Finance ☐ Industries and manufacturing ☐ Technology ☐ Marketing and Advertising ☐ Communication ☐ Entertainment and Media ☐ Retails ☐ Services and E-commerce ☐ Other ☐ None
Which type of work?	☐ Manual work ☐ Office work ☐ Both ☐None

[1] This refers to areas characterized by higher population density and extensive human-built environments such as cities or towns. Urban areas typically have various amenities, services, and infrastructure. Typically, the population living in towns of 2,000 people or more or in national and provincial capitals is classified as urban.

[2] This pertains to areas with lower population density and less built-up infrastructure. Rural areas often have more open spaces, agricultural lands, and natural landscapes compared to urban areas.

[3] Main work is considered as the work that takes most of your time in general.

Lifestyle (mark with an x)	
Are you smoking cigarettes?	☐ Yes(currently) ☐ Yes(previously) ☐ No
On average, how many cigarette packs[1] do/did you smoke per day?	☐ None ☐ Less than 1 ☐ 1 to 2 ☐ More than 2
How long have you been smoking? (Write a number in years[2])	
Are you regularly exposed to secondhand smoke[3]?	☐ Yes ☐ No
How many days per week do you consume alcohol[4]?	☐ 0 ☐ 1 to 3 ☐ 4 to 6 ☐ 7
How many standard alcoholic drinks[4] do you consume per week?	☐ None ☐ 1 to 2 ☐ 3 to 4 ☐ 5 to 7 ☐ More than 7
How would you describe your current level of physical activity?	☐ Sedentary (less than 30 min of moderate physical activity per week) ☐ Little active (30 to 89 min of moderate physical activity per week) ☐ Moderately active (90 to 149 min of moderate physical activity per week) ☐ Active (150 min of moderate physical activity or more per week)
On average, how many minutes per day do you engage in physical activity?	☐ Less than 15 minutes ☐ 15 to 29 minutes ☐ 30 to 60 minutes ☐ More than 60 minutes
On average, how many days per week do you engage in physical activity?	☐ Less than 2 days ☐ 2 to 4 days ☐ More than 4 days
How long do your typical physical activity sessions last? (Write in minutes)	
What is your average daily sedentary time[5]?	☐ Less than 5 hours ☐ 5 to 10 hours ☐ More than 10 hours
How many hours per day do you engage in prolonged sitting[6]?	☐ Less than 2 hours ☐ More than 2 hours

[1] Suppose that 1 pack includes 20 cigarettes.
[2] Put zero (0) if you have never smoked and one (1) if you smoke less than 1 year.
[3] Daily exposure to the tobacco smoke of others at home, work, or public places
[4] Frequency of drinking alcoholic beverages (e.g., 354 ml can/bottle of beer, 118ml glass of wine, 44ml shot of hard liquor
[5]Sedentary time refers to periods when an individual engages in very low physical activity or movement (sitting at work, at school, at home, in a car/bus/train, and during leisure time (e.g., watching TV, playing video games, using the computer, reading, socializing)). It includes any time spent in activities with minimal energy expenditure.
[6]Prolonged sitting specifically refers to extended periods of sitting without breaks or movement. It highlights the negative effects of sitting for long stretches without interruptions or physical activity.

Diet (mark with an x)	
How often do you consume fruits[1] and vegetables[1] in your meals?	☐ Daily ☐ Several times a week[2] ☐ About once a week ☐ Rarely[3] ☐ Never
How often do you eat processed meat (sausages, bacon, etc.)?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you include low-fat dairy products in your diet?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you consume white meat, such as poultry or fish?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you eat whole grains[4]?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you consume sugary drinks[5]?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you consume sugary desserts[6]?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never
How often do you eat fast food[7]?	☐ Daily ☐ Several times a week ☐ About once a week ☐ Rarely ☐ Never

[1] Examples of fruit: fresh fruit, chopped, cooked or canned fruit, dried fruit, fruit juice. Examples of vegetables are raw leafy vegetables, cooked, canned, frozen, or chopped vegetables, and vegetable juice.
[2] Several times a week: This means that you consume the item more than once in a week, but not every day. It indicates a frequency that is more than occasional but less than daily.
[3] Rarely: This means you consume the item infrequently, on special occasions, or very seldom. It indicates that the item is not a regular part of your diet.
[4] Whole grain is defined as cooked brown rice or other cooked grain, cooked 100% whole-grain pasta, cooked hot cereal, such as oatmeal, uncooked whole grain pasta, brown rice or other grain, 100% whole grain bread, 100% whole grain muffin, 100% whole grain ready-to-eat cereal
[5]Sugary drinks such as soft drinks (excluding diet soda), vitamin drinks, energy drinks, and speciality coffee with syrup (e.g., mocha)
[6]Desserts containing sugar, such as candy, chocolate bars, cake, cookies, and ice cream
[7]Includes foods from fast food restaurants (e.g., burgers, fries, tacos), pizza, and instant meals (e.g., instant ramen noodles)

Dietary Supplements[1] (mark with an x)	
Do you take dietary supplements1?	☐ Yes ☐ No
How often do you consume omega 3? (Answer based on the previous question)	☐ Never2 ☐ Rarely3 ☐ Often4
How often do you consume multivitamins? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume vitamin B6? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume vitamin C? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume vitamin D? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume magnesium? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume calcium? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume iron? (Answer based on a previous question)	☐ Never ☐ Rarely ☐ Often
How often do you consume probiotics[5]?	☐ Never ☐ Rarely ☐ Often
How often do you consume fibre supplements?	☐ Never ☐ Rarely ☐ Often
How often do you consume folic acid (females only)?	☐ Never ☐ Rarely ☐ Often
Please write the names of any other supplements you take.	

[1]Dietary supplements are intended to add to or supplement the diet and are different from conventional food.

[2]Never: Indicates that the supplement is not consumed at all.

[3]Rarely: Indicates that the supplement is consumed infrequently or occasionally but not on a regular basis.

[4]Often: Indicates that the supplement is consumed frequently or regularly as part of the dietary routine.

[5] Probiotics are a combination of live beneficial bacteria and/or yeasts.

Stress-PSS4 (mark with an x)	
In the last 2 months, how often have you felt that you were unable to control the important things in your life?	☐ Never ☐ Almost Never ☐Sometimes ☐ Fairly Often ☐ Very Often
In the last 2 months, how often have you felt confident about your ability to handle your personal problems?	☐ Never ☐ Almost Never ☐Sometimes ☐ Fairly Often ☐ Very Often
In the last 2 months, how often have you felt nervous and stressed?	☐ Never ☐ Almost Never ☐Sometimes ☐ Fairly Often ☐ Very Often

ANNEX No. 3. SAMPLE COLLECTION & MANAGEMENT

1. SCOPE OF THE PROCEDURE

This SOP describes the processes for collecting biological samples from study participants and managing the collected samples from the Clinical Partners and Test facility. Specifically, it provides instructions for:

a. the collection, labelling, storage, and shipment of biological samples from the Clinical Partners to the Test Facility.
b. The receipt, inspection, handling, storage, recording, archiving, and disposal of biological samples by the Test Facility.

2. DEFINITIONS

- **Samples**: serum & plasma samples collected from subjects enrolled in the study.
- **Test Facility:** the partner that performs the biological analysis of samples. Protavio Ltd (PAO) (former Protatonce Ltd) is the Test Facility for the DIOPTRA project.
- **Collection tubes**: serum/plasma tubes used for initial blood collection prior to centrifugation.
- **Transfer tubes:** 15mL centrifuge tubes are used to transfer the upper liquid phase (serum or plasma samples) after centrifugation.
- **Storage tubes**: 2mL microcentrifuge screw-cap tubes that are used to aliquot and store samples.

3. EQUIPMENT / MATERIALS

3.1 EQUIPMENT

#	Description	Specifications	Recommended Cat No
1	Centrifuge	1300-1800 g (RCF) 18-25 °C For 16mm x 100mm tubes	N/A
2	Ultra-low Freezer	-80 °C or below	N/A
3	Pipette	Single channel 200-1000uL range	Rainin Pipet-Lite LTS Pipette L-1000XLS+, #17014382

Funded by the European Union

4	Laminar flow hood (optional)	Class II A2 cabinet	N/A
5	Racks for collection/transfer/storage tubes	See tube specifications	VWR, # 211-0204 (for 2mL tubes)
6	Personal Protective Equipment	Lab coat, gloves, etc	N/A

3.2 MATERIALS

#	Description	Specifications	Recommended Cat No
1	9-10mL Serum collection tubes	Plastic, 16x100mm, with clot activator (silica), **red cap colour**, transparent	BD Vacutainer, #367896 Greiner Vacuette, #455092
2	9-10mL K2EDTA Plasma collection tubes	Plastic, 16x100mm, with K2EDTA additive, **purple/lavender cap colour**, transparent	BD Vacutainer, #367525 Greiner Vacuette, #455045
3	15mL centrifuge tubes (transfer tubes)	nonpyrogenic and DNase-/RNase-free	Corning, #430791
4	2mL screw-cap microcentrifuge tubes (storage tubes)	nonpyrogenic and DNase-/RNase-free, non-sterile, freezable to -80 °C, can be centrifuged to 12,000×g, with silicone O-ring screw-caps	VWR, #525-0651 (tubes), #525-0653 (screw-caps)
5	Cryoboxes with dividers, 9x9 positions	133x133x50mm size, resistant to temperatures down to –140 °C, standard waterproof coating	VWR, #479-1417 (boxes), #479-1465 (dividers)

4. IDENTIFICATION

4.1 DIOPTRA PARTICIPANT ID

DIOPTA Participant IDs are aimed to differentiate participants and to ensure the anonymization of personal data during the submission of samples to the Test Facility performing the biological analysis.

The following identification system **should be followed** for the codification of participants:
Each ID will include the clinical site code followed by a 5-digit number that is unique to each participant and follows a continuous numbering starting from 0001. Continuous numbering is based on the date of the participant's inclusion in the study (date of signature of informed consent).

Clinical Site Codes:

Clinical Site	Clinical Site Code
KBDZ	CP09
CHUL	CP12
RSYD	CP15
UKCM	CP16
BURGOS	CP21
NKUA	CP01
GOC	CP23
AG.SAVVAS	CP24

An example of a DIOPTRA Participant ID is CP24-0034.

4.2 SAMPLE ID

Sample IDs are aimed to differentiate samples and aliquots coming from the same participant:
The following nomenclature is proposed for Sample IDs:
For serum samples:
DIOPTRA ID-S-N, where S stands for serum, and N is the number of aliquots*.
Example: CP24-0034-S-1

For plasma samples:
DIOPTRA ID-P-N, where P stands for plasma and N is the number of aliquots*.
Example: CP24-0034-P-1

*Aliquot numbering is optional in case the clinical partner needs to catalogue every tube collected in their database management system.

4.3 TUBE LABELING

Each storage tube should be clearly labelled either using digital labels or handwritten with permanent ink.
Each tube stored and shipped to the Test Facility should contain the following information:
- Sample ID
- Collection Date

5. PROCEDURE FOR SERUM & PLASMA COLLECTION

Serum and Plasma samples will be collected from each subject following a blood draw.
NOTES:
- The blood draw should be performed before the colonoscopy.
- First, draw blood for serum, then draw blood for plasma.
- Follow best practices to avoid hemolysis of samples.

1. First, fill in the name and signature of the responsible clinical partner in the Sample Collection Form
2. Fill in the details of the participant and the sample collection date.

5.1 SERUM

1. Draw blood into one serum collection tube (red capped). Record the time of the blood draw in the Collection Form.
2. Gently invert the tube 5-6 times to mix blood with the clot activator.
3. Place upright on a test rack and allow to sit for 30-60 min at 18-25ºC until clotting has occurred.
4. Centrifuge at 1,500-2,000 x g for 10 minutes at 18-25ºC. Record the time of initiation of centrifugation in the Collection Form.
5. Using a pipette, collect the upper liquid phase (serum) into a 15mL transfer tube, taking care not to remove any of the clotted material.
6. Prepare 4 aliquots of 500µL using the 2mL storage tubes. Use correctly labelled tubes (see section 4.0 IDENTIFICATION). Record the number of aliquots prepared for DIOPTRA in the Collection Form.

Note: Left-over serum samples can be kept for internal biobanking by the clinical partner. Leftover samples can be handled according to clinical partner internal procedures.

7. Store serum aliquots in cryoboxes in an ultra-low freezer at -80ºC or below.

5.2 PLASMA

1. Draw blood into one K2EDTA collection tube (purple capped). Record the time of blood draw in the Collection Form.
2. Gently invert the EDTA tube 8-10 times immediately after the blood sample has been taken to avoid microclotting.
3. Centrifuge immediately (or within 1 hr from blood draw) at 1,500-2,000 x g for 10 minutes at 18-25ºC.
4. Using a pipette, collect the upper liquid phase (plasma) into a 15mL transfer tube, taking care not to remove any of the middle and lower layers containing blood cells.
5. Prepare 4 aliquots of 500µL using the 2mL storage tubes. Use correctly labelled tubes (see section 4.0 IDENTIFICATION). Record the number of aliquots prepared for DIOPTRA in the Collection Form.

Note: Left-over plasma samples can be kept for internal biobanking by the clinical partner. Left-over samples can be handled according to clinical partner internal procedures.

6. Store plasma aliquots in cryo boxes in an ultra-low freezer at -80ºC or below.

5.3 COLONOSCOPY PROCEDURE

During the colonoscopy, the details related to the procedure and the quality of the procedure should be entered into the Sample Collection Form as presented in the form.

6. SHIPMENT TO TEST FACILITY

6.1 NUMBER OF ALIQUOTS TO BE SHIPPED

Each clinical partner should submit to the Test Facility:

- 2 x 500µL serum aliquots per patient &
- 2 x 500µL plasma aliquots per patient

Note: The remaining 2 aliquots of each type per patient should be kept by the Clinical Partner as reserved back-up material.

6.2 PACKAGING

The aliquots should be placed in cryoboxes and a map of the position of the aliquots corresponding to each patient ID in the box should be provided to the Test Facility by the Clinical Partner in excel format. Each cryobox should also be numerically labelled to avoid confusion during receipt.

6.3 PERIODICITY OF SHIPMENTS

Shipments should be arranged every 6 months by the clinical partner.

6.4 SHIPPING INSTRUCTIONS

1. Samples should be shipped in **dry ice** with **next-day courier delivery services**.

Note: Do NOT use FedEx, as this courier does not deliver dry ice to Greece.

2. Use the following sample description:

UN3373 Biological Substance Cat B packed in Dry Ice, Class 9, UN1845 ….. kgs. Use for research purposes only.

3. Ensure cryoboxes are fully covered with dry ice during transport.
4. Arrange shipment between **Monday and Wednesday** to ensure that the package is delivered by the end of the week.

Shipping Address
Protavio Ltd
NCSR Demokritos
Lefkippos Technology Park, Bldg 27
Patriarchou Grigoriou E' & 27 Neapoleos Str.
15341, Ag. Paraskevi
Attiki
Greece

Contact Person:
Nikos Tsolakos
Email: nikos.tsolakos@protavio.com
Tel: +30 210 9610307

7. TEST FACILITY RECEIPT & TEMPORARY STORAGE

Biological Samples should be received by trained personnel, and the Biological Sample Receipt Form should be completed, dated and signed.

Trained personnel should provide a general description of received samples, including:

a. Number of Biological Samples (boxes, tubes etc)
b. Quantity (volume) per tube (approximate)
c. Identification numbers
d. Shipping Temperature

All relevant documentation that accompanies the shipment should be retained and handed to the Study Director.

In case of deviations from the packing list (i.e. different number of boxes or vials received) or shipping temperature (i.e. shipment not in dry ice or samples appear defrosted due to lack of dry ice), these deviations should be recorded in the relevant Receipt form.

Upon receipt, biological samples should be immediately stored at -80°C.

8. TEST FACILITY INSPECTION & STORAGE

Biological Samples should be inspected by the Study Director to ensure that the correct Biological Samples have been received, under the correct conditions and that they are uniquely identified and recorded.

The Study Director needs to perform the following activities:

- Verify the identity of Biological Samples. Verification should include ensuring that information on the container in which the test item is shipped and the labelling on the test item matches information recorded by the organisation responsible for the study on accompanying documentation and study protocols.
- Check that the types of samples, number of tubes and quantities are correct based on accompanying documentation and study protocols.
- Check the physical characteristics of the Biological Samples match the expected characteristics. Specifically, serum and plasma samples should be in liquid form (frozen) and appear yellow. Any deviations, e.g. hemolytic samples, samples received defrosted etc., should be recorded.
- Check transportation documents (including the Biological Sample receipt form) for correct shipment conditions.

Observations, including deviations, should be recorded in the Biological Sample Receipt Form_Study Director.

Upon inspection, biological samples should be continuously stored at -80^{o}C. If samples are expected to be thawed multiple times, they should be further aliquoted in smaller volumes.

9. TEST FACILITY RECORDS

Biological Samples should be recorded in a Biological Sample Inventory. It is the responsibility of the Study Director and the Test facility to maintain, amend and archive this Inventory. The Inventory should contain, at minimum, the following information:

- Sample ID
- DIOPTRA Participant ID
- Sample Type
- Collection date
- Sender (Clinical Partner)
- Lab Reception Date
- Date of reception by Protavio (PAO)
- No tubes received
- Volume per tube
- Total volume received
- Visual Inspection results (normal, hemolytic, icteric or lipemic)
- Storage Temperature
- Storage Location
- Box ID
- Comments/Deviations from the receipt process
- Comments/Deviations during sample collection (upon inspection of Sample Collection Form by the clinical partner)

10. CLINICAL CHARACTERIZATION

Funded by the European Union

The clinical information of each participant will be collected in the Medical Information/History Case Form and will be used to categorize samples into the four groups of the clinical protocol. Collection of Medical Information per participant is performed by the Clinical Partners and is uploaded to the DIOPTRA prospective platform.

11. ARCHIVING

All documentation related to the receipt, storage and inventory of biological samples will be archived by the Test Facility.

The Test Facility will receive two aliquots of serum and two aliquots of plasma samples from each Clinical Partner per participant. Each Clinical Partner will retain two aliquots of serum and two aliquots of plasma samples per participant as backup material for the DIOPTRA study, along with any left-over sample volumes for internal biobanking.

DIOPTRA samples will be retained for the duration of the DIOPTRA project.

12. DISPOSAL

Biological samples will be disposed of at the end of the archival period and destroyed by incineration following internal disposal protocols and dedicated services for the disposal of biological hazardous material.

ANNEX No. 4. COLONOSCOPY AND SAMPLE COLLECTION CASE FORM

SAMPLE COLLECTION FORM

DIOPTRA

(BASED ON SOP279-01.Sample Collection & Management)

CLINICAL PARTNER INFORMATION	
CLINICAL PARTNER	
COLLECTION RESPONSIBLE FULL NAME	
COLLECTION RESPONSIBLE SIGNATURE	

PARTICIPANT GENERAL INFORMATION	
PARTICIPANT FULL NAME	
DIOPTRA ID	
SAMPLE COLLECTION DATE	

COLONOSCOPY PROCEDURE INFORMATION	
Sedation Drug	
Sedation Drug Dose	
Use of CO_2	☐ YES ☐ NO
BBPS right colon (0-3)	☐ 0 ☐ 1 ☐ 2 ☐ 3
BBPS transversum (0-3)	☐ 0 ☐ 1 ☐ 2 ☐ 3
BBPS left colon (0-3)	☐ 0 ☐ 1 ☐ 2 ☐ 3
BBPS overall (0-9)	☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9
Cecal intubation	☐ YES ☐ NO
Time to cecal intubation (min)	

Funded by the European Union

Withdrawal time (min)	
Time required for interventions (min)	
Total Procedure Time (min)	
Gloucester Comfort Score (1-5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

SERUM COLLECTION		
Collection Tube Lot #		
Collection Tube expiration date (YYYY/MM)		
Time of blood draw (HH: MM)		
Time centrifugation was initiated (HH: MM) No earlier than 30 min from the blood draw		
Total Waiting Time from blood draw till centrifugation initiated (min) No longer than 60 min		
Number of storage tubes prepared (0.5 mL per tube)		
Left-over volume (mL) (if applicable)		
Storage Temperature (^{O}C)		
Visual Inspection	☐ Normal	☐ Hemolyzed (red colour)
	☐ Icteric (bright yellow)	☐ Lipemic (turbid)
Comments – Deviations from SOP Different Tube types, longer waiting times, low sample volumes, other centrifugation conditions, etc.		

PLASMA COLLECTION	
Collection Tube Lot #	
Collection Tube expiration date (YYYY/MM)	
Time of blood draw (HH: MM)	
Time centrifugation was initiated (HH: MM)	

Total Waiting Time from blood draw till centrifugation initiated (min) No longer than 60 min		
Number of storage tubes prepared (0.5 mL per tube)		
Left-over volume (mL)		
Storage Temperature (^{o}C)		
Visual Inspection	☐ Normal	☐ Hemolyzed (red colour)
	☐ Icteric (bright yellow)	☐ Lipemic (turbid)
Comments – Deviations from SOP Different Tube types, longer waiting times, low sample volumes, other centrifugation conditions, etc		

ANNEX No. 5. MEDICAL INFORMATION/HISTORY CASE FORM

MEDICAL INFORMATION/HISTORY CASE FORM

Family history (mark with an x)	
Do you have any family history of CRC?	☐ 1st-degree relatives (Parents, Children and Siblings) ☐ 2nd-degree relatives (Grandparents, Grandchildren, Aunts, Uncles, Nephews, Nieces and half-Siblings)" ☐ Both ☐ No
What was the age of the youngest relative at diagnosis?	☐ Less than 50 years ☐ More than 50 years ☐ Unknown ☐ No family history
What is the sex of the youngest relative?	☐ Male ☐ Female ☐ No family history

Personal History (mark with an x)	
How many colonoscopy(ies) did you have in the past?	☐ 0 ☐ 1 ☐ 2 ☐ More than 2
How many years ago was the last one? (Put zero (0) for no previous colonoscopy)	
What were the findings of your last colonoscopy?	☐ Healthy ☐ Non advanced adenoma ☐ Advanced adenoma ☐ No previous colonoscopy
Do you have Diabetes Type II?	☐ Yes ☐ No
If applicable, what is the potential measurement of your HbA1c[1] (glycated haemoglobin)?	☐ Less than 5.7% ☐ 5.7% to 6.4% ☐ More than 6.5%
Do you have hypertension (high blood pressure)?	☐ Yes ☐ No
Do you have dyslipidemia (high levels of fat – cholesterol and triglycerides in the blood)	☐ Yes ☐ No
Do you have cardiovascular disease?	☐ Yes ☐ No
Do you have chronic kidney disease?	☐ Yes ☐ No
Do you have any allergies or asthma?	☐ Allergy☐ Asthma ☐ Both ☐ None

If yes, what are you allergic to?	
At what age were you diagnosed with allergies?	☐ Less than 10 years ☐ 10 to 19 years ☐ More than 20 years ☐ No allergies
At what age were you diagnosed with asthma?	☐ Less than 10 years ☐ 10 to 19 years ☐ More than 20 years ☐ No asthma

[1] Bennett, C. M., Guo, M., & Dharmage, S. C. (2007). HbA1c as a screening tool for detection of type 2 diabetes: a systematic review. *Diabetic medicine*, *24*(4), 333-343.

Medication (mark with an x)	
Have you taken any medication within the last month?	☐ Antihypertensives ☐ Anticoagulants ☐ Aspirin ☐ "Non-steroidal anti-inflammatory drugs (NSAIDs) ☐ Glucagon-like peptide-1 (GLP-1) ☐ Sodium-Glucose Transport Protein 2 SGLT-2 ☐ Dipeptidyl Peptidase-4 (DPP-4) ☐ Statin ☐ Insulin ☐ GLP-1 ☐ SGLT-2 ☐ Metformin ☐ Antiplatelet ☐ Corticosteroids ☐ DPP-4 ☐ Other ☐ No
If applicable, please specify any other medication you are currently taking.	

Symptoms (mark with an x)	
Are you experiencing abdominal pain?	☐ Yes ☐ No
Have you noticed a change in your defecation habits?	☐ Diarrhea ☐ Constipation ☐ Both ☐ No
Have you observed blood in the stool?	☐ Yes ☐ No
Are you experiencing bleeding from the rectum?	☐ Yes ☐ No
Are you experiencing symptoms such as gas, abdominal cramps and/or bloating?	☐ Yes ☐ No
Do you feel that your rectum is not completely empty after having a bowel movement?	☐ Yes ☐ No
Have you been diagnosed with anaemia?	☐ Yes ☐ No

Females only

How many pregnancies have you had?	☐ 0 ☐ 1 to 2 ☐ More than 3
At what age did you have your first pregnancy?	☐ Before 30 years old ☐ After 30 years old ☐ No pregnancy
Have you ever used oral contraceptives?	☐ No ☐ Yes (previously) ☐ Yes (currently)
What is your current menopausal status?	☐ None ☐ Pre-menopausal ☐ Menopausal ☐ Post-menopausal

ANNEX No. 6. DIOPTRA SOFTWARE COMPONENTS

A. DIOPTRA Software Frontend - Anonymization Tool (EHR data)

The Anonymization Tool, developed by CSCY and deployed at each clinical site, provides an extra layer of privacy protection to already pseudonymised structured medical data. It is used for EHR data prior to uploading to the Data Software Backend. The tool employs the k-anonymity method with the Mondrian algorithm on the input data, which was developed due to the possibility of indirect identification of records from public databases. Users can upload a file (following the DIOPTRA data template of T3.1), define variables to be anonymised, set the k parameter for k-anonymity, run the algorithm, and download the resulting file with the anonymised dataset.

B. DIOPTRA Software Frontend - Clinical Dashboard

The Clinical Dashboard developed by CSCY for uploading, deleting, or reviewing clinical data allows clinical partners' staff to upload their anonymised datasets on the DIOPTRA platform, receive feedback (successful uploading or error reports about inconsistencies with the data template), get informed about the volume and quality of uploaded data, and delete records. The functionalities of the dashboard interfaces and access rights were defined and implemented as described in deliverable D2.1, with some specific points mainly affecting the prospective study to be incorporated in a later version of the clinical dashboard. The Use-Cases (UC) that are implemented in the Clinical Dashboard are:

- UC1 - Prospective / Retrospective Data Acquisition and Upload from Clinical Sites
- UC2 - Delete Action of a Data Record from the Clinical Side
- UC3 - Uploaded Data Overview
- UC4 - Data Review for Single Participant
- UC6 - Follow-up Study Monitoring (later versions)

Figure 1 shows the sign-in and upload screens of the DIOPTRA clinical dashboard, while Figure 2 shows the screen for the uploaded data overview.

Funded by the European Union

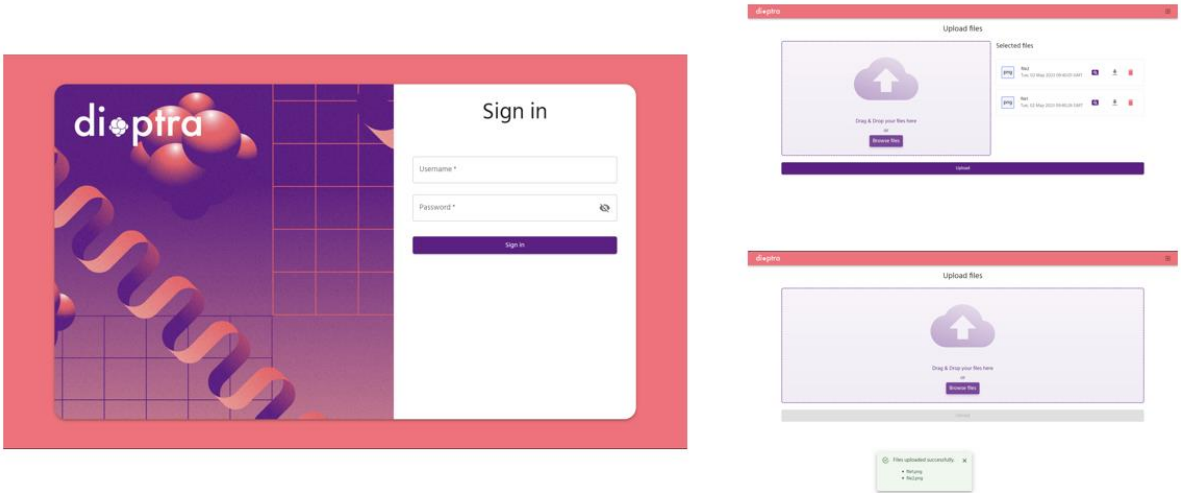

Figure 3. Sign in and upload screens to the DIOPTRA clinical dashboard

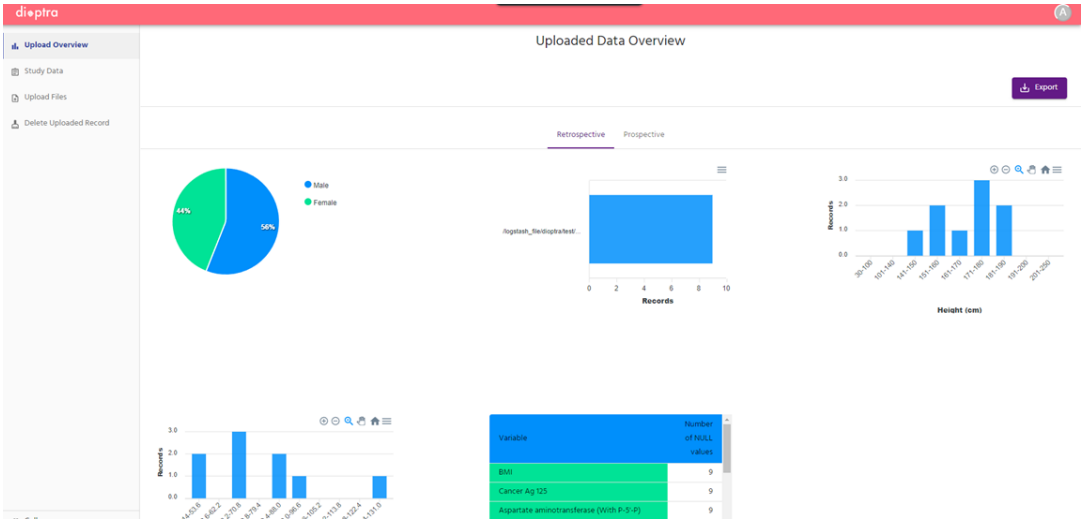

Figure 4. Uploaded data overview

C. DIOPTRA Software Backend

The anonymised retrospective and prospective data will be stored in the DIOPTRA Software Backend by utilising the ELK STACK. The DIOPTRA Software Backend developed by INTRA consists of several key components. RESTful APIs are developed to expose the curated datasets to the project's assets, with access regulated using the Keycloak service. Logstash pipelines were generated to handle data ingestion, with separate data indices assigned to each unique clinical site. Data format and compliance with the curation rules are ensured at the pipeline level. To control data access and client communications, a Keycloak instance is provided as an OpenID Connect (OIDC) service. User accounts

have been created for all project clinicians, with four distinct roles (Clinical Manager, Clinical Staff, IT Administration Staff, and IT Staff), and users are grouped by their clinical site.

Five unique curation categories are applied to the datasets, with distinct curation rules created in collaboration with the clinical partners and incorporated through the Logstash pipelines:

- Header Row Structure
- Field Format
- Variable value ranges
- Correlation-based rules
- Missing Values

Meaningful messages are sent for display in the Clinical Dashboard in case of errors, describing the error type and the row it's located in.

The DIOPTRA Software Backend also comprises a Programming Environment by CSCY, which provides access to a defined sub-dataset and tools that could be utilised for analysis and/or pattern recognition and model development.

Mobile Application

The mobile application, among other functionality (knowledge base, recommendations, etc.), collects and uploads behavioural questionnaire data from follow-up study participants. It consists of three discrete components: i) the mobile app itself, which will run on users' smartphones, ii) the mobile app backend, which is a server-side component responsible for data management, user authentication, the core business logic, the API endpoints, and the push notifications, and iii) the Content Management System (CMS), based on a web interface, which is an internal tool that interacts with the mobile app backend and allows easy management and publishing of content that will be present on the app's health literacy module. The mobile app will be available through Google and Apple app stores. Formal applications will be sent for approval by the App stores. In case of approval delay or decline, the app will be accessible through smartphone browsers. The terms of use of the mobile app have been prepared, and the mobile app will be offered to users under these terms. Each participant enrolled in the DIOPTRA study will need to input their unique DIOPTRA ID to store their responses in the database. The participant must complete a questionnaire based on sociodemographics, lifestyle, diet, supplement consumption, and stress categories. Specific suggestions promoting a healthy lifestyle, encouraging the adoption of healthier eating habits, aligning with the direction of limiting alcohol consumption, promoting smoking cessation, and encouraging physical fitness will be triggered based on the user's

responses. The pseudonymised questionnaire data, along with the provided suggestions, will be stored in the DIOPTRA central storage platform and will be available for clinical users to review via the Questionnaire Dashboard. User manuals for the mobile application will be developed to ensure that participants can effectively navigate and utilise the application's features.

Overall, the mobile app is structured around four primary features:

a. Behavioural questionnaires: Participants enrolled in the DIOPTRA study will be requested to complete up to two behavioural questionnaires depending on their recruitment phase: the baseline behavioural questionnaire (for all participants) and the follow-up questionnaire (for follow-up study participants). Initially, the baseline questionnaire will be provided on paper at the clinical site. Since the follow-up study's mobile application will only engage healthy and non-AA groups, and diagnoses won't be available during the baseline phase, the paper format will be used universally at this stage. Then, the app will be introduced to follow-up study participants, who will be requested to retake the behavioural questionnaire in order to identify post-baseline changes and developments.
b. The Health Literacy Module (HLM): Its purpose is to educate the public about CRC and associated health behaviours. This module will offer a collection of resources, including articles, infographics, and videos, all curated by medical professionals to ensure the information is accurate and current. Topics will cover CRC risk factors, colonoscopy guidelines, and detailed explanations of procedures, presented in a user-friendly format.
c. Diary: The diary feature is a user-centric functionality designed to encourage and track individual lifestyle behaviours as well as mild bowel symptoms (such as chronic constipation, chronic diarrhoea, etc.) and identify any changes or recent trends in those topics. To promote user engagement and ensure that it is used by the user, it will be presented in the form of a periodic set of questions that the user can answer before continuing to use the app. The user will receive notifications on their phone whenever new questions are available for them to answer.
d. RiskAssessment Module (RAM) .

The Risk Assessment Module (RAM) is designed to utilise both prospective and follow-up data, seamlessly integrating into the mobile application. It will analyse behavioural data derived from the behavioural questionnaire (Annex 2), along with clinical information from the medical case form questionnaire (Annex 5), by employing an AI-based model to address a multi-classification problem. The RAM will predict a participant's DIOPTRA group categorisation based on their behavioural profile. The module will output probability scores indicating the category to which a participant is most likely

to belong. Based on these probabilities scores, the wellness levels will be categorised into low, medium and high. Additionally, a ranking of the risk factors can be produced, providing a score for each factor based on the overall wellness score. The RAM will then generate tailored suggestions and recommendations accordingly. These recommendations will be sourced from well-known public organisations and articles and validated through the Health Literacy Module (HLM), where clinical experts will approve their relevance and efficiency. The RAM will be designed to continuously monitor and update the assessment of the wellness score based on new data inputs. As participants provide additional data through the mobile application periodically, the RAM will refine its predictions and recommendations, ensuring they remain relevant and accurate over time. Beyond generating wellness scores and recommendations, the RAM can be utilised to design personalised health interventions. By understanding a participant's specific risk factors, tailored intervention programs can be developed to address these behavioural risks proactively, potentially reducing the likelihood of developing CRC based on behavioural factors. Effectively, the RAM is set to function as a recommender system, as provided in Figure 3, providing personalised guidance based on the participant's behavioural data.

Funded by the European Union

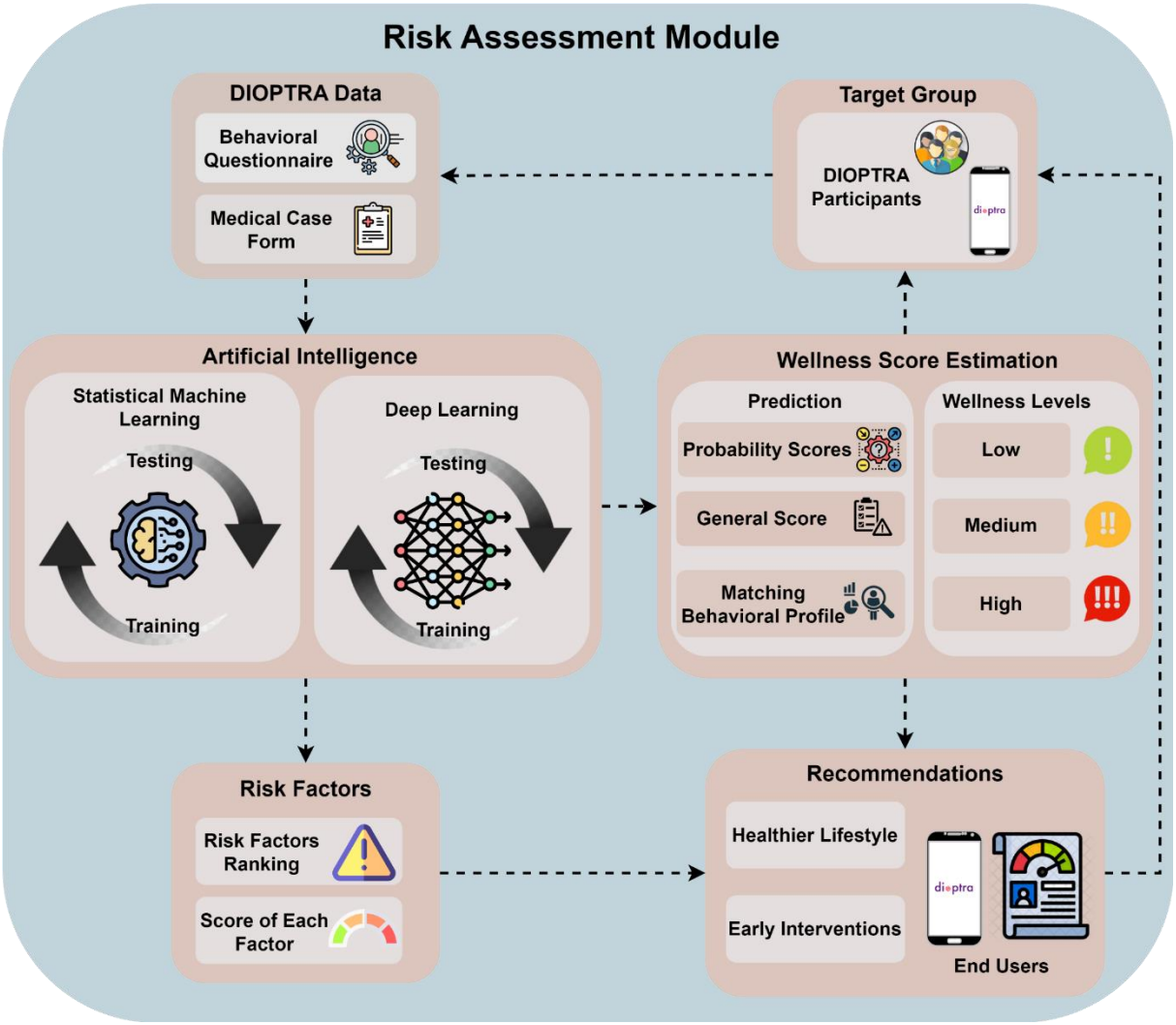

Figure 5. The Risk Assessment Module Architecture

Regarding the communication between the RAM and the overall DIOPTRA system, Figure 4 illustrates the data flow from the completion of the questionnaires at the clinical sites to the recommendations provided in the mobile application. The process begins with the DIOPTRA participant filling out the questionnaires on paper during their visit to the hospital. Subsequently, the clinician encodes the participant's responses into the data template, and the anonymisation tool is applied to ensure data privacy. A unique DIOPTRA ID is then assigned to the participant, and the file is uploaded to the DIOPTRA software. Participants enrolled in the follow-up study will be notified to download the mobile application and sign in using their corresponding DIOPTRA ID. The mobile application will retrieve the participant's responses from the DIOPTRA software and transmit them directly to the RAM. The RAM will process the data, and the results, including the wellness score, specific recommendations, and risk factor rankings, will be displayed in the mobile application. These results will also be sent back to the DIOPTRA software for further analysis and record-keeping.

Funded by the European Union

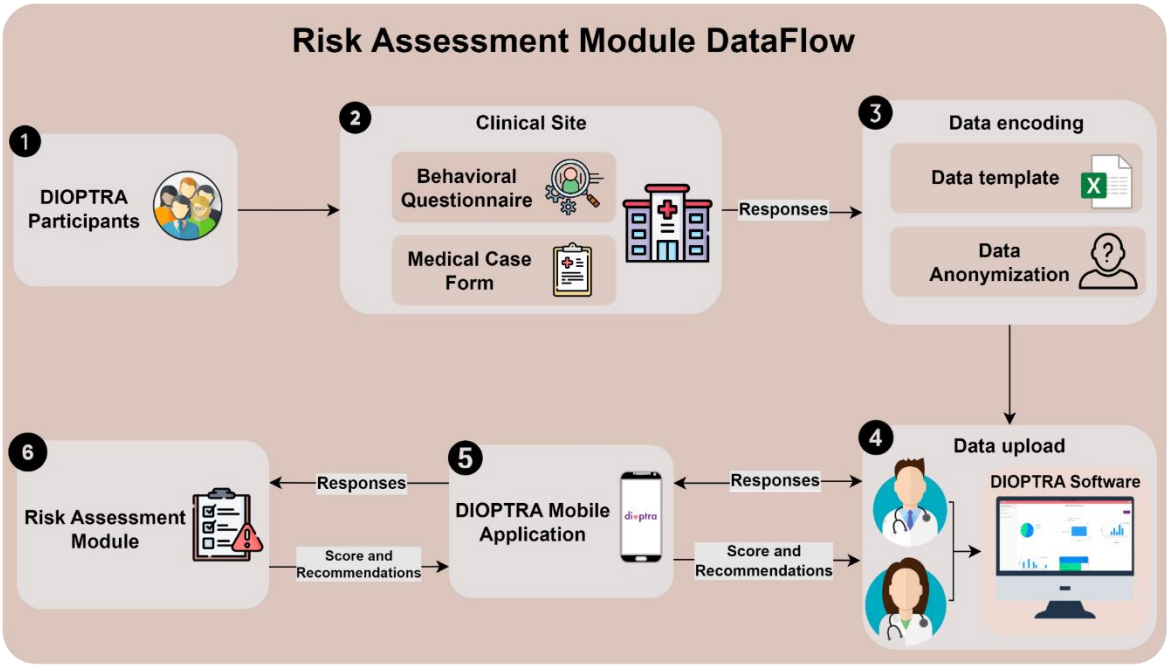

Figure 6. RAM Data Flow and communication with the DIOPTRA system

Figure 5 provides a detailed illustration of the process by which behavioral questions are presented to users of the mobile application, as well as how the wellness score and personalized recommendations are provided.

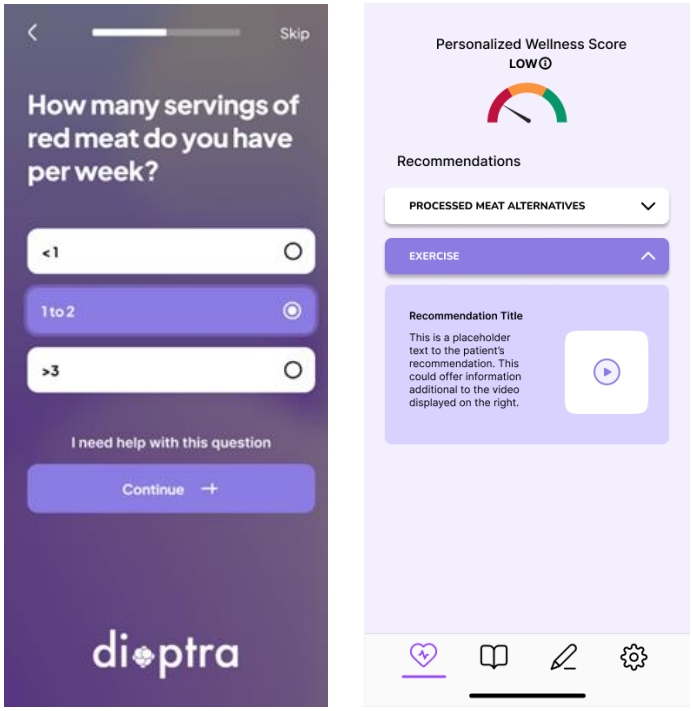

Figure 7. Sample screens of the mobile app

Risk Factor Analysis

The Risk Factor Analysis (RFA) aims to identify and evaluate the association of various risk factors with CRC and to predict early CRC risk. Advanced AI techniques, combined with statistical analysis, will be employed to assess the significance of each identified risk factor. This approach will enable the ranking of the most critical factors influencing CRC, thereby providing insights for early detection and prevention strategies. The collected prospective data will be analysed using state-of-the-art AI models, including Machine Learning (ML) and Deep Learning (DL) algorithms. These models will be trained to recognise complex patterns and interactions among the risk factors, enhancing the predictive accuracy of CRC risk assessments. Statistical techniques such as regression analysis, correlation studies, and factor analysis will be utilised to quantify the importance of each risk factor. These techniques will help determine the relative contribution of each factor to the overall risk of developing CRC. The analysis will consider a wide range of variables, including demographic data, lifestyle factors, medical history, and clinical biomarkers. Visualisation tools will be employed to increase the transparency and interpretability of the developed models. These tools will generate plots and charts that illustrate the relationships between risk factors and CRC. Visualisations such as heatmaps, correlation matrices, and

importance ranking plots will provide a clear understanding of the data, enabling researchers and clinicians to detect patterns and trends effectively. The architecture of the RFA is presented in Figure 6.

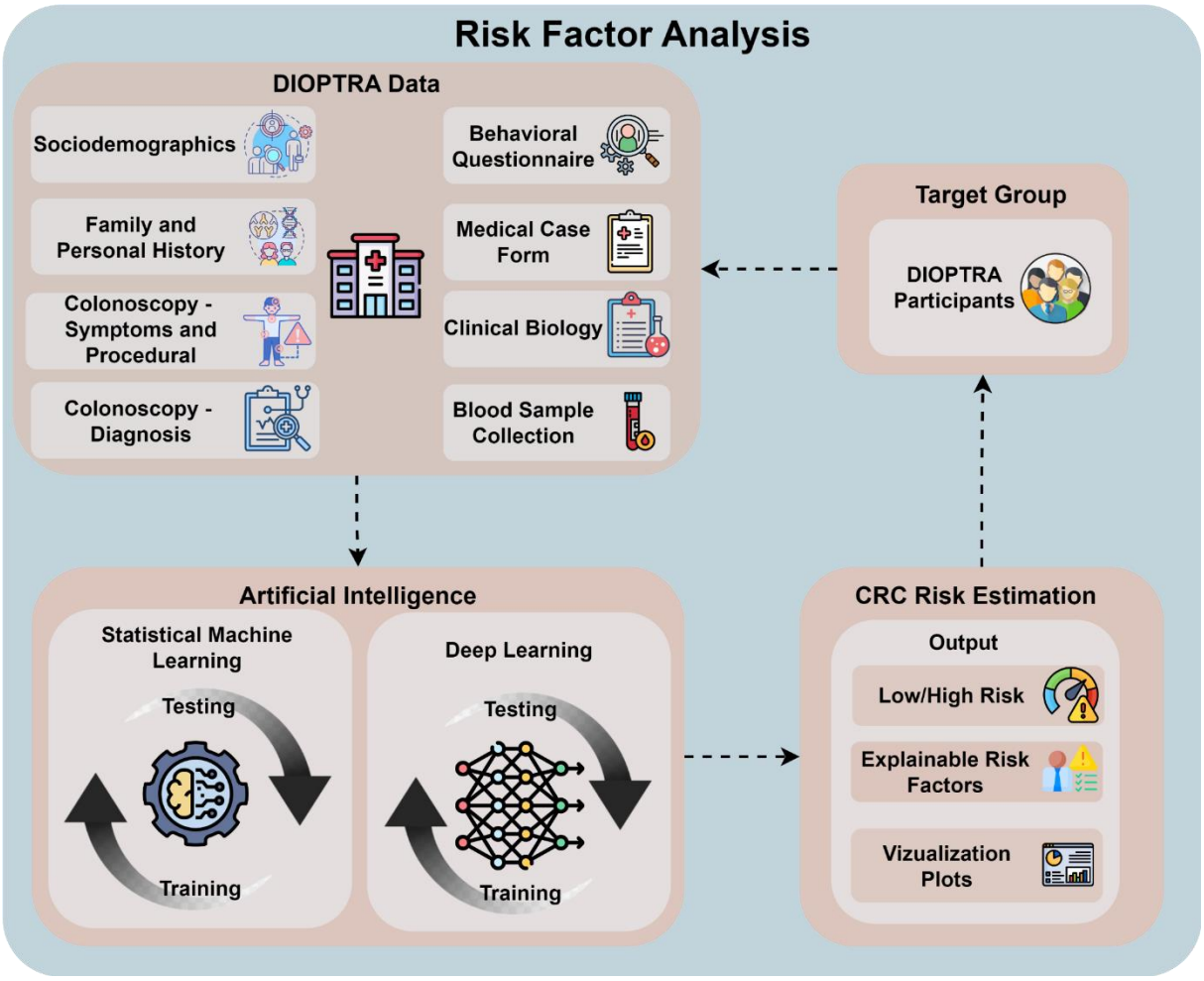

Figure 8. Architecture of the RFA

Regarding the communication between the Risk Factor Analysis (RFA) module and the overall DIOPTRA system, Figure 7 illustrates the data flow from the visit of participants to hospitals to the identification of risk factors. During these hospital visits, a wide range of data will be collected from participants, including sociodemographic information, family and personal medical history, responses to the behavioural questionnaire, medical case form data, colonoscopy symptoms and procedural details, clinical biology from blood samples, information related to blood sample collection, and colonoscopy diagnosis results. Clinicians will encode all this information into the standardised data template. Once the data is encoded, the file will be uploaded to the DIOPTRA software. Through retrieval mechanisms, the uploaded data will be utilised to identify risk factors and examine their association with (CRC).

Funded by the European Union

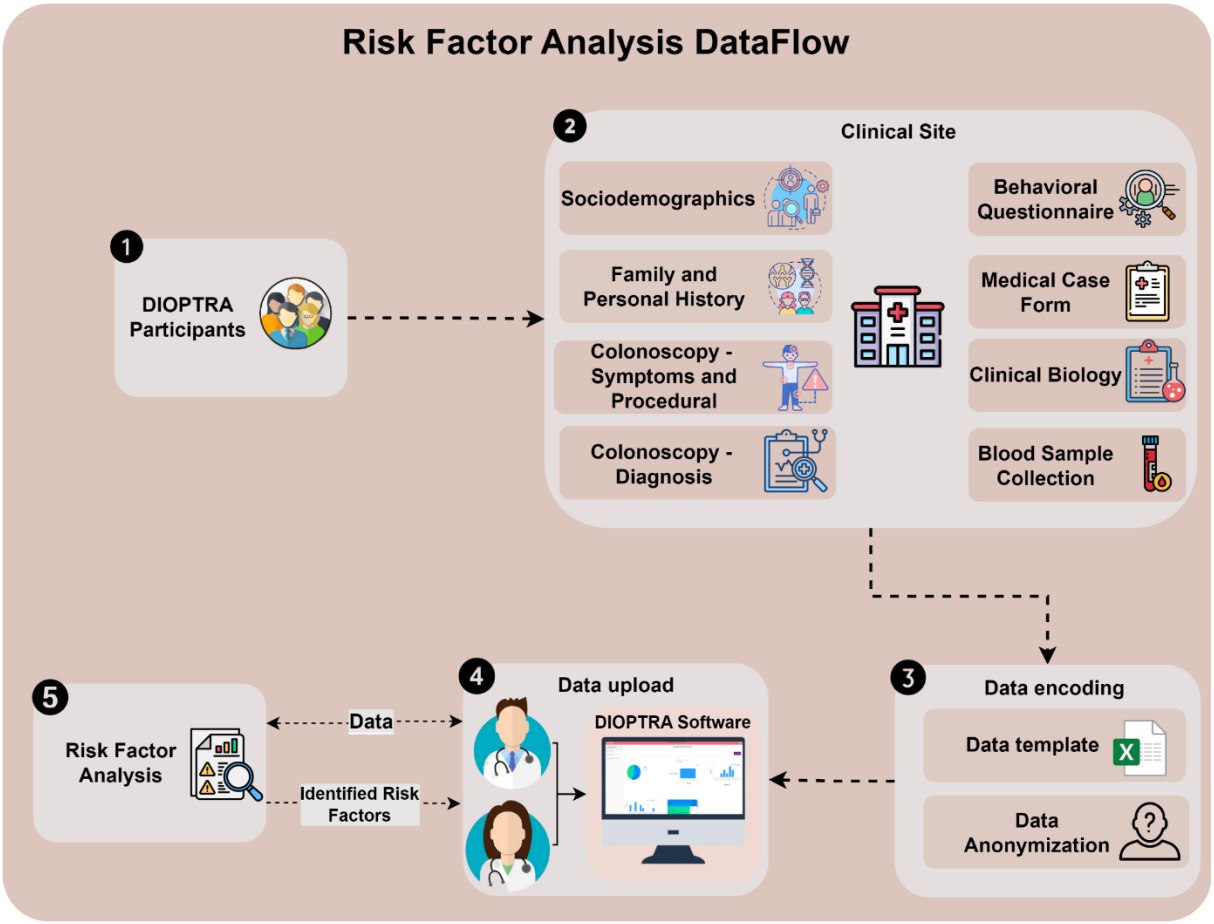

Figure 9. RFA Data Flow and communication with the DIOPTRA system

AI Modelling

This component leverages advanced Artificial Intelligence (AI) techniques to classify individuals into four distinct categories: Healthy, Non-advanced adenomas, Advanced adenomas, and Colorectal Cancer (CRC). By utilising protein concentration data and various risk factors from the prospective study, the component employs machine learning algorithms to identify significant biomarkers. The AI models will be trained, validated, and cross-validated to ensure robustness and accuracy, with a focus on achieving over 85% CRC screening specificity. Additionally, this component is designed to enhance diagnostic precision and provide actionable insights. Therefore, it emphasises explainability, ensuring that the AI's decision-making process is transparent and interpretable, potentially recommending further medical procedures such as colonoscopy based on the identified risk profile.

The AI component retrieves data from the DIOPTRA backend. The first step is data preprocessing, which generates preprocessed data used to train models. The validated models are stored in the model

repository. The validated models and the test data pass to the risk prediction step that extracts the results, which are sent back to the DIOPTRA backend. The explanation module provides explanations based on the transparency information(input data & predicted output) from the risk prediction module.

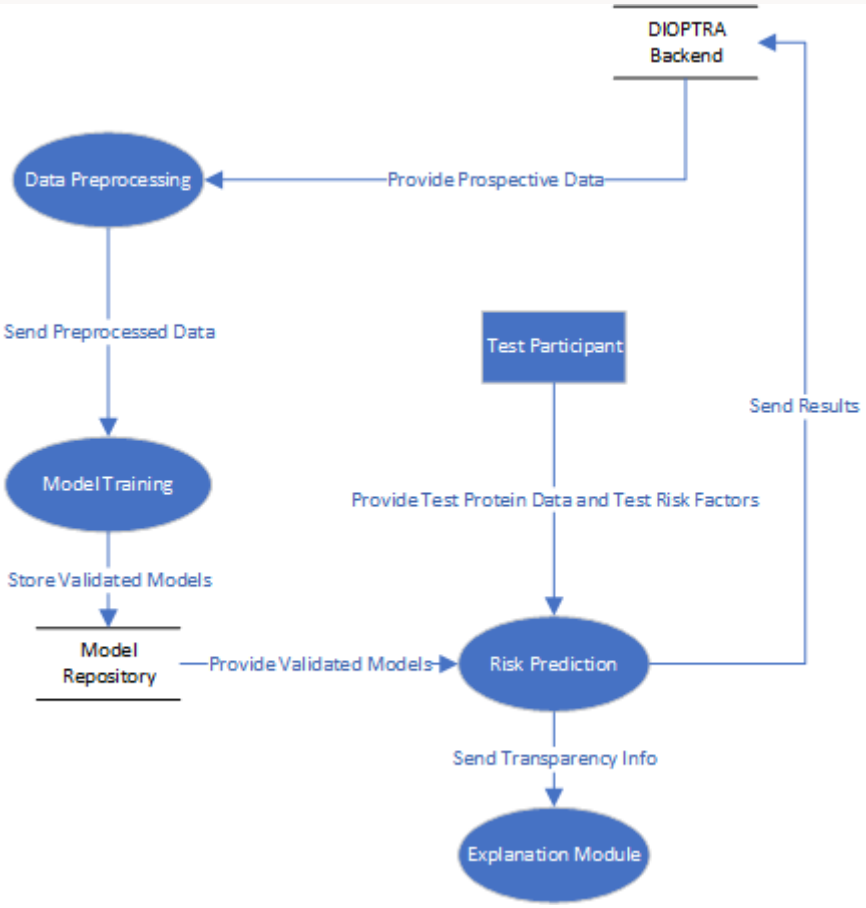

Figure 10. AI Data Flow and communication with the DIOPTRA system

Hosting Infrastructure

The project will use the GRNET infrastructure for data storage. GRNET S.A. is a public sector technology company in Greece that has been operating since 1998, providing networking, cloud computing, HPC, data management services, and e-infrastructures to academic and research institutions, educational bodies, and public sector agencies operating under the auspices of the Ministry of Digital Governance. In the context of the DIOPTRA Horizon project, GRNET will provide us with the following infrastructure and equipment, namely specific virtual machines (VMs):

- **VM1: Master Node, Logstash, Kibana, API Gateway**
  8 or 16 cores, 32 GB RAM, 500 GB disk (pref SSD)

- **VM2: Elastic Data nodes 1 & 2**
  8 cores, 16 GB RAM, 200 GB disk
- **VM3: Elastic Data nodes 3 & 4**
  8 cores, 16 GB RAM, 200 GB disk
- **VM4: Programming Environment & Interface**
  4 cores, 8 GB RAM, 100 GB disk
- **VM5: Staging Environment for all Services**
  8 cores, 16 GB RAM, 100 GB disk
- **OS: CentOS Linux**