D6.2: DIOPTRA Clinical Protocols

Revision: v.1.0

Work package	WP6
Task	T6.1
Due date	31/08/2023
Submission date	31/08/2023
Deliverable lead	TERAGLOBUS
Version	1.0
Authors	Rūta Goštautaitė (TERABLOBUS)
Reviewers	Christos Fotis (PAO) Zheshen Jiang (CHUL)
Abstract	This deliverable presents protocols for the retrospective, discovery, and prospective clinical studies of DIOPTRA that have already been or will be approved by national bioethics committees. The protocols contain detailed descriptions of procedures, patient samples, applied research methods, and questionnaires for patients and medical personnel.
Keywords	Retrospective clinical study, Prospective clinical study, Biomarker discovery study, Colorectal Cancer Screening

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	07/08/2023	1st version	Rūta Goštautaitė (TERAGLOBUS), Laura Baronienė (TERAGLOBUS), Gintarė Marinė (TERAGLOBUS), Christos Androutsos (UOI), Grigoris Antonopoulos (INTRA), Christos Fotis (PAO), Dimitra Stefanatou (ARTHUR), Toygar Oruc (ARTHUR), Marcela Chavez (CHUL), Zheshen Jiang (CHUL), Sabrina Kral (GRAZ)

Grant Agreement No.: 101096649
Call: HORIZON-MISS-2021-CANCER-02
Topic: HORIZON-MISS-2021-CANCER-02-01
Type of action: HORIZON-RIA

V0.2	30/08/2023	2nd version including comments from doctors	Rūta Goštautaitė (TERAGLOBUS), Catherine Loly (CHUL), Sisse Njor (RM-RRH), Ioannis Papanikolaou (NKUA)
V0.3	31/08/2023	Pre-final version based on feedback from internal reviewers	Rūta Goštautaitė (TERAGLOBUS), Zheshen Jiang (CHUL), Christos Fotis (PAO)
V1.0	31/08/2023	Deliverable approved by the PC and submitted	Stavros Miloulis (ICCS)

DISCLAIMER

Funded by the European Union

Funded by the European Union (DIOPTRA, 101096649). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

Project funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**DEM**	
Dissemination Level		
PU	*Public, fully open, e.g., web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✓
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patent filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

This deliverable focuses on the retrospective and prospective studies' clinical protocols used within the DIOPTRA project across the participating clinical sites. It serves as a guiding framework for the consortium's clinical partners involved in the project's execution.

- The retrospective study aims to create a consolidated dataset from Electronic Health Records, focusing on demographic, medical, and family history data. AI models will analyse these data to identify risk factors for early CRC prediction.

- The DIOPTRA discovery study aims to discover a panel of blood-based diagnostic protein biomarkers with a verified connection to the colorectal cancer mechanism for CRC screening and early detection.

- The prospective study seeks to validate the DIOPTRA screening system clinically. Its primary objective is to establish the sensitivity and specificity of the system in detecting CRC, validated against clinical diagnoses (colonoscopy).

Overall, DIOPTRA aims to advance Colorectal Cancer (CRC) understanding, diagnosis, and prevention through innovative technologies, Artificial Intelligence (AI), and data-driven approaches, ultimately enhancing healthcare outcomes and resource efficiency.

Funded by the European Union

TABLE OF CONTENTS

DISCLAIMER **2**
Copyright notice 2
EXECUTIVE SUMMARY **4**
TABLE OF CONTENTS **5**
LIST OF FIGURES **8**
ABBREVIATIONS **9**
1 INTRODUCTION **10**
1.1 Purpose of the Document 11
1.2 Structure of the Document 11
2 LITERATURE ANALYSIS AND RATIONALE **13**
2.1 Literature Analysis 13
2.2 Rationale of the Study 18
3 RETROSPECTIVE CLINICAL STUDY PROTOCOL **20**
Document History 21
3.1 Synopsis 21
3.2 General Information 23
3.2.1 Rationale for the Retrospective Study 23
3.2.2 Background 24
3.3 Study Plan 25
3.3.1 Study Hypothesis 25
3.3.2 Study Design 25
3.3.3 Number of Investigation Sites and Study Duration 28
3.4 Study Methods 29
3.4.1 Data Collection Requirements 29
3.4.2 Source Documents 29
3.4.3 Study Deviations and Clinical Study Protocol Changes 29
3.5 Quality Control Procedures 30
3.5.1 Data review and processing 30
3.5.2 Study Suspension or Early Termination 30
3.5.3 Study Close Out 30
3.6 Data Reporting and Publication 31

Funded by the European Union

3.6.1 Data Reporting and Publication 31
3.7 Risks and Benefits 31
3.7.1 Risk-to-Benefit Rationale 31
3.8 Bibliography 32
4 BIOMARKER DISCOVERY STUDY PROTOCOL 39
4.1 Medical University of GRAZ Project Partners 40
4.2 Title, Acronym, Unique Identifier 40
4.3 Cooperation Partners 41
4.4 Background 42
4.5 Objectives 43
4.6 Hypothesis 44
4.7 Sample Selection 44
4.8 Inclusion Criteria 45
4.9 Exclusion Criteria 45
4.10 Study Design: Study Type 45
4.11 Study Design: Data Collection 46
4.12 Study Design: Material Collection 47
4.13 Study Design: Methods and Data Analysis 47
4.14 Study Design: Archiving 49
4.15 Study Design: Disposal 50
4.16 Statistical Analysis 50
4.17 Ethical Considerations 51
4.18 Data Processing and Access 52
4.19 Study Duration and Funding 52
4.20 Annex 52
4.21 Literature 52
5 PROSPECTIVE CLINICAL STUDY PROTOCOL 54
5.1 Document History 56
5.2 Abbreviations and Acronyms 56
5.3 Synopsis 57
5.4 Background and Rationale 59
5.5 Study Design 60
5.6 Objectives 62

Funded by the European Union

5.7 Endpoints 62
5.8 Study Population 62
5.9 Sample Size 64
5.10 Study Procedures 66
5.10.1 Overview 66
5.10.2 Enrolment 67
5.10.3 Blood Sample Collection 67
5.10.4 Colonoscopy and Diagnosis 67
5.10.5 Collection of DIOPTRA Data 67
5.10.6 End of Study 67
5.10.7 Follow-up Study Timing and Procedures 68
5.10.8 Data Flow and Data Processing (prospective, follow-up) 69
5.10.9 Data Analysis 73
5.11 Ethics and Data Management 74
5.12 Quality Control Procedures 74
5.13 Bibliography 78
ANNEX No. 1. Informed Consent Form 88
ANNEX No. 2. Behavioural Questionnaire 99
ANNEX No. 3. Samle Collection & Management 104
ANNEX No. 4. Colonoscopy and Sample Collection Case Form 111
ANNEX No. 5. Medical Information/History Case Form 114
6 CONCLUSIONS 116
REFERENCES 117

LIST OF FIGURES

FIGURE 1: DIOPTRA CLINICAL TIMELINE 10

ABBREVIATIONS

AI	Artificial Intelligence
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
BLOCKS	Blocks Health and Social Care EOOD
BURGOS	Fundacion Burgos Por La Investigacion De La Salud
CHUL	Centre Hospitalier Universitaire De Liege
CIP	Clinical Investigation Protocol
CRC	Colorectal Cancer
FU	Follow Up
GOC	Linac-Pet Scan Opco Limited
NKUA	National and Kapodistrian University of Athens
PSD	Prospective Study Design
RM-RRH	Region Midtjylland
SOP	Standard Operating Procedure
UKCM	Univerzitetni Klinicni Center Maribor

Funded by the European Union

1 INTRODUCTION

This Deliverable outlines the retrospective and prospective study protocols of the DIOPTRA project, focusing on advancing the understanding and application of Colorectal Cancer (CRC) screening, diagnosis, and treatment, leveraging innovative technologies, Artificial Intelligence (AI), and data-driven approaches. DIOPTRA seeks to facilitate the development of new products, services, and policies, ultimately aiming at better healthcare outcomes, efficient resource management, and stronger public-private-people partnerships. A key objective of DIOPTRA is to create a routine blood test for CRC screening that's universally accessible, catering to all age groups, and aimed at detecting individuals who might not otherwise undergo screening based on prevailing European or national guidelines.

Protocols for the retrospective and prospective clinical studies of DIOPTRA have already been or will be approved by national bioethics committees. The protocols contain detailed descriptions of retrospective and prospective clinical studies, patient samples, applied research methods, and questionnaires for patients and medical personnel. The report serves as a key reference for all consortium partners and stakeholders involved in the project.

The detailed clinical timeline that will be followed throughout the implementation of the DIOPTRA studies and the use of the data collected and generated is depicted in the following Figure:

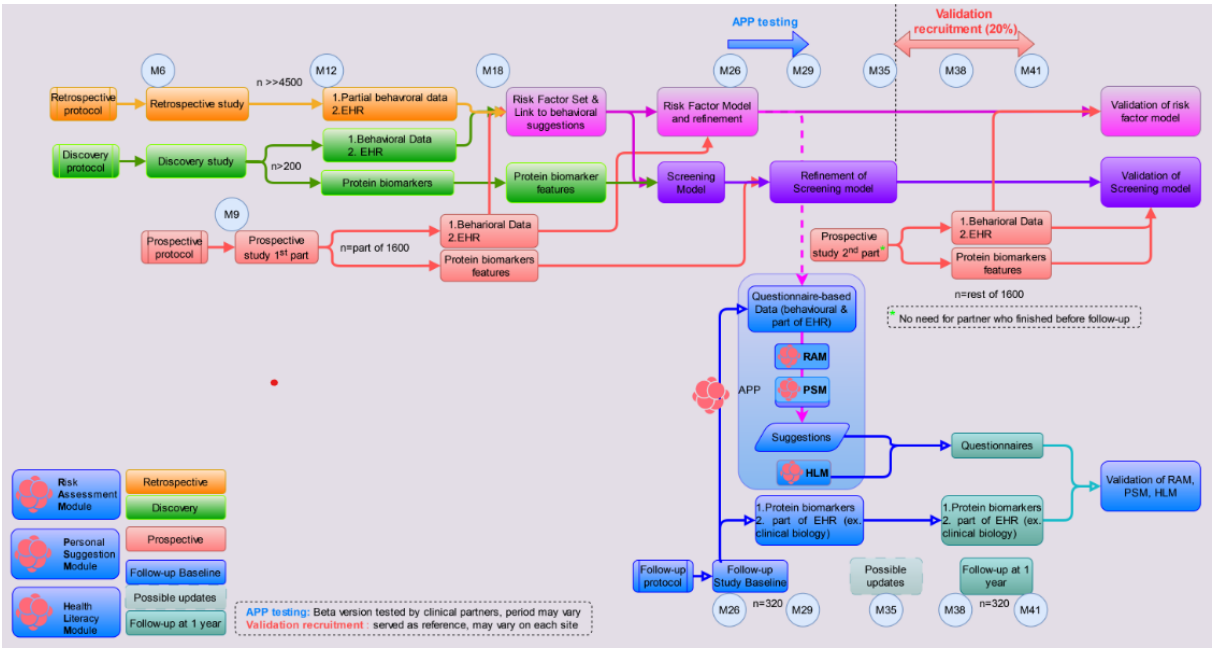


Figure 1: DIOPTRA Clinical Timeline

More information with regard to the specific components and actions will be described in the dedicated Deliverable on the full DIOPTRA specifications (D2.1).

1.1 PURPOSE OF THE DOCUMENT

The purpose of this document is to present the protocols of the retrospective and prospective clinical studies for the implementation of the DIOPTRA project in all participating clinical sites. Some information on the clinical studies is provided below:

- The retrospective study aims to generate a homogenised dataset consisting of retrospective data on demographic, medical, and family history information derived from the Electronic Health Records of the clinical sites for the risk factor investigation on early Colorectal Cancer (CRC) risk. The main objective of the retrospective study is to analyse all the risk factors and to determine the best features as input to the AI models for predicting early CRC risk. These factors can also generate suggestions for daily habits that may be very useful in preventing colorectal cancer.

- The discovery study aims to identify a panel of blood-based diagnostic protein biomarkers with a verified connection to the colorectal cancer mechanism for CRC screening and early detection. This part of the DIOPTRA project will identify a panel of blood-based protein diagnostic biomarkers that can differentiate between the DIOPTRA study groups (healthy individuals, individuals with non-advanced adenomas, individuals with advanced adenomas, and individuals with early-stage CRC).

- The purpose of the prospective clinical study is the clinical refinement and validation of the DIOPTRA screening system. The main objective is to validate the diagnostic sensitivity and specificity for CRC detection of the DIOPTRA screening system using clinical diagnosis as a reference (colonoscopy). Another aim of the study is to validate some of the risk factors identified in the retrospective study. Overall, DIOPTRA seeks to utilise the final group of proteins within blood tests that will be able to identify citizens who should undergo further colonoscopy screening. The validation of the biomarker signature has many advantages since it is: (1) obtained through a minimally invasive technique (blood sampling), (2) inexpensive, and (3) 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.

1.2 STRUCTURE OF THE DOCUMENT

This document is structured into the following main sections for clarity and ease of reference:

1. **Introduction**: This section provides an overview of the document, outlining its purpose and structure.

2. **Literature analysis and Rationale**: This section discusses the global CRC landscape, current diagnostic methods, and the significance of DIOPTRA's contributions to CRC diagnosis, research, and prevention.

3. **Retrospective clinical study protocol**: The protocol submitted by the clinical sites that obtained bioethical permits in their countries is presented, outlining the methodology for the retrospective study.

Funded by the European Union

4. **Biomarker discovery study protocol:** Retrospective biomarker discovery pilot study created and carried out by one partner (GRAZ).

5. **Prospective clinical study protocol**: This section presents the detailed clinical study protocol to be submitted for bioethical approval by the clinical sites.

6. **Conclusions**

Funded by the European Union

2 LITERATURE ANALYSIS AND RATIONALE

2.1 LITERATURE ANALYSIS

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 tumors 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 estimations 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

In the CRC treatment domain, options include endoscopic and surgical excision, radiotherapy, immunotherapy, palliative chemotherapy, targeted therapy, extensive surgery, and local ablative therapies for metastases [1]. Meanwhile, screening methods consisting of endoscopic tests (e.g., colonoscopy) and non-invasive alternatives such as the fecal immunochemical test (FIT) have been put into action. Applied pathways have successfully inhibited cancer progression [4], decreasing mortality rates through 2017. Moreover, some EU countries have adopted population-based screening programs over the last 15 years, seeking to halt incidence and mortality rates. In this regard, 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 and 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 broad 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 pre-malignant states.

Pressing Conditions

In determining the CRC risk status, factors such as age, BMI, diet, smoking habits, and family history [8] have been pinpointed by researchers and clinicians alike. Namely, age, sex, and family history have been integrated into practice as flagships on risk stratification [2]. However, evidence on the complete risk factor set has yet to be analysed in the context of a detailed assessment. Indicatively, Western registry data show an increased incidence in the age group of 40-44, considerably lower than the 50-year threshold [9]. This tendency is attributed to modern lifestyle alterations, although more assays are required as to the corresponding effects. Despite the long-assumed CRC preventability based on modifiable risk factors, awareness and knowledge exploitation still need to be higher. Moreover, despite the available arsenal of screening practices, citizen participation could be improved due to suboptimal performance and the invasive or overall taxing nature besetting

Funded by the European Union

these methods [10]. EU reports indicate participation rates of 14%, a somewhat disappointing number compared to the >60% rate for breast cancer screening programs [11]. Poor screening outreach is augmented by limited penetration of Council recommendations into clinical practice. By 2019, only three member states had adopted population-based screening targeting all at-risk citizens, albeit purely based on age thresholds of 50-74 years. Over the last three years, other states have also launched population-based screening or regional programs. Still, there needs to be more progress on standardised programs that unite knowledge on CRC towards EU-wide regulations. 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. In this vein, liquid biopsy is a promising new tool for non-invasive, quick, and safe assessment [6]. However, lack of operating protocols, reproducibility issues, and cost-effectiveness discrepancies are impeding clinical implementation. 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 needs 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.

Liquid Biopsy CRC Biomarkers

Advances made since the human genome was first sequenced have greatly increased cancer genetics knowledge and contributed to the development of cutting-edge molecular biology tools. Several genetic and epigenetic alterations have been found to initiate and sustain specific deregulated cellular signaling pathways involved in CRC tumors. Today, liquid biopsy, as a novel tool, can easily detect these alterations in body fluids without investigating the initial tumor site [12]. Liquid biopsy is non-invasive (or minimally invasive), fast, and easy to perform without requiring sophisticated equipment or highly skilled human intervention. Some companies provide kits for extracting stool samples at home, thus speeding up the clinical process. The result can be obtained in a few hours, compared to tissue extraction and pathological analysis, which may require several days. Overall, the field of CRC biomarker extraction from blood samples for screening purposes has been enriched with numerous studies demonstrating a high TRL, justifying DIOPTRA's ambition for formulating a discriminative protein subset. For instance, a sensitive blood-derived CRC screening test using the methylated Septin 9 (SEPT9) biomarker has been developed for detecting most CRCs in all stages. The test can be conducted for average-risk individuals unwilling or unable to undergo colonoscopy [21]. In a pivotal prospective trial including almost 8,000 subjects, methylated SEPT9 detected colorectal cancer with 48% sensitivity and 92% specificity. The sensitivity of this test decreased to 35% for stage I disease, 63% for stage II disease, 46% for stage III disease, and 11% for advanced adenoma. The overall sensitivity for detecting colorectal cancer was superior to gFOBT testing but less than FIT's. An assay for SEPT9 is commercially available and FDA-approved for individuals refusing other screening tests, although it has not been incorporated into official guidelines. Similarly, the anti-p53 antibody represents another promising biomarker with consistent findings in one screening cohort and the 3–4 years before diagnosis in two prospective cohort studies [22]. Meta-analysis data indicated that serum p53 antibody possesses potential diagnostic value for CRC. However, discrimination power was somewhat limited due to the low sensitivity [23]. Another biomarker investigated is carcinoembryonic antigen (CEA) [24], a complex glycoprotein produced by 90% of colorectal cancers and contributes to the malignant characteristics of a tumor. It can be quantitively measured in serum samples, with its plasma level constituting an exploitable disease

marker. However, because of its lack of sensitivity in the early CRC stages, CEA measurement is considered unsuitable for population-based screening. In addition to CEA, the dynamic measurements of carbohydrate antigen 19-9 (CA19-9) and carbohydrate antigen 125 (CA125) are believed to monitor the prognosis of colorectal cancer patients. CEA and CA19-9 are useful in CRC patients' preoperative, postoperative prognosis, and follow-up. In the field of more newly studied biomarkers, blood serum macrophage inhibitory cytokine-1 (MIC1) could serve as a novel diagnostic marker of early-stage colorectal cancer [25]. MIC1 is markedly upregulated in colorectal malignancies and premalignant adenomas, and its serum levels are strongly associated with neoplastic progression within the large bowel [26]. Therefore, MIC1 could be a candidate complementary biomarker for screening early-stage CRC by combination with CEA [27]. Likewise, Galectin-3 expression was correlated with the genesis and development of colorectal cancer. It could be used as a biological marker for prognosis, present at levels 30-fold higher in CRC patients than normal controls. Galectin-3 ligand is stable for long periods in blood, and the assay can be performed on volumes of 5μL or less, differentiating healthy from malignant samples (irrespective of cancer stage) and advanced adenomas with blinded verification across sample sets. Finally, ESM1, CTHRC1, and AZGP1 are the additional highly reliable and easily detectable biomarkers capable of secreting into blood, urine, and saliva by integrating transcriptomics and proteomics data at the system biology level [28]. DNA-based biomarkers have also been investigated, with cell-free circulating DNA (cfDNA) showing prospects for identifying residual diseases and assessing treatment response and prognosis [6]. Maintaining the same genomic signatures with the matching tumor tissue, cfDNA allows for quantitative and qualitative evaluation. More specifically, PCR-based approaches (COLD-PCR, PNAs-LNA, ARMS, etc.), digital PCR (ddPCR and BEAMing), NGS (deep sequencing, TAM-seq, Safe-Seqs, CAPP-Seq, cSMART, digital sequencing), SERSnanotags and UltraSEEK comprise the major cfDNA techniques used in CRC studies. However, when it comes to cost-effective CRC screening, there exists a clear cost-sensitivity trade-off, with the cost-effective DNA-based biomarkers carrying low sensitivity, hindering their capacity for feasibly substituting for current front-line screening techniques such as gFOBT and FIT. Moreover, these techniques require pre-determination of the examined genes, which demands a rigorous development phase for bridging the gap "from lab to market." Although individual marker genes have been measured with a statistically significant presence in CRC patients, most are primarily detectable in advanced or metastatic cases, with few (like NEUROG1) bearing prominence for more early stages [29]. With recent advances in sequencing analytics, micro-RNA markers extracted from whole-blood samples are gradually gaining intense research interest [30]. Studies have shown remarkably high diagnostic sensitivity, though results are pending corroboration and thus immature for transition to population-based screening soon. On another note, one of the key prospects of the above biomarkers in CRC (or other cancer types) analysis corresponds to their exploitability beyond population-based screening in investigating metastasised malignancies stemming from an unknown primary cause. Such is the case for Cancer of Unknown Primary (CUP), a term applied to a heterogeneous group of metastatic tumors, the primary sites that cannot be identified at the time of diagnosis despite extensive investigations [31]. The history of CUP is characterised by early metastasis from the unknown primary site, aggressive course, and resistance to conventional chemotherapy. Unfortunately, the processes by which this orphan disease originates and progresses have not been fully elucidated, and its biology remains unclear. Advances in cancer detection techniques have contributed to reducing CUP incidence from around 3–5% in the 1990s to 1–2% in the current era [31], while – as with initial CRC screening – research is gradually following the liquid biopsy biomarker path to remove the "U" out of "CUP" by efficiently identifying the primary malignancy, grant the opportunity for site-specific treatment that has the potential to improve projected outcomes. However, task requirements and parameters differ from the initial screening mission, with investigated serum CRC biomarkers like CA19-9 and CA125 bearing

Funded by the European Union

poor predictive value for CUP cases [31]. That is expected since many studies found negative colonoscopies characterised by several CRC-predicted CUP cases. On the other hand, molecular gene profiling has yielded an identification rate of 77–94% using second-generation microRNA-based assays, gene expression profiling-based microarray tests, or quantitative-PCR low-density arrays. The most well-known marker for metastatic carcinomas of gastrointestinal (GI) origin is CDX2 (Caudal Type Homeobox 2), a nuclear transcription factor contributing to intestinal epithelial cell proliferation and differentiation [32]. Virtually all colorectal adenocarcinomas are positive for CDX2, which is also expressed in neuroendocrine tumors of the GI tract. However, the evidence base is insufficient to recommend routine molecular screening [33]. Therefore, to this day, PET/CT scans remain the state of the art in clinical practice by a favorable cost-effectiveness ratio, bearing a sensitivity of 87% and specificity of 71% [31]. Under the described circumstances, key open research questions remain about the identification of sensitive biomarkers among the vast range of available options. Both initial screening (appealing to the general population) and screening for primary (appealing to CUP patients) comprise clinical application fields that a corresponding breakthrough could immediately benefit. Of note, optimal results should not be expected to pinpoint any single biomarker, as studies on limited sets have demonstrated the superior sensitivity and specificity performance of collective subsets as opposed to individual markers [34].

Artificial Intelligence for Cancer Screening

Under DIOPTRA architecture, AI will be exploited in the context of a) blood biomarker analysis for identifying a discriminative subset, b) assessing an extended risk factor pool compared to current limited clinical consideration, and c) clinical decision support for stratifying high-risk cases as an output from our front-line screening tool to the main evaluation phase. As such, AI has been widely employed in biomarker evaluation, from drug development to pathology and oncology [35]. Generally, outcome prediction has been the hallmark of AI utilisation in cancer. Indicatively, AI models have attempted to predict 5-year CRC recurrence risk, outperforming grading and/or staging evaluation by expert pathologists. As regards screening, AI has been implemented for automated decision-making in combination with various studied screening procedures [7]. Relying on the duration of the gradual transition path from normal mucosa to a premalignant growth and then to a malignant lesion spanning over 15-20 years, AI can detect suspect changes corresponding to abnormal tissue, which may be indicative of either a premalignant precursor lesion or an early-stage tumor. A high adenoma detection rate (ADR) has been validated as inversely correlated with adenoma miss rate and the risk of post-colonoscopy CRC), with every 1% increase in ADR corresponding to a 3% reduction in CRC development risk and a 5% reduction in fatal CRC incidence. However, ADRs may range from 7% to 53% among different endoscopists, creating the demand for objective and reproducible assessment towards attaining a robust ADR in clinical settings. In this regard, convolutional neural networks (CNNs) have been found to detect and localise premalignant lesions on imaging data accurately. In a prospective randomised controlled trial [36] with conventional vs. computer-aided colonoscopy detection, ADR was significantly increased in favor of computer-aided detection (CAD). Similarly, virtual colonoscopy paired with machine learning (ML) modules could distinguish between benign and precancerous colorectal polyps in an average-risk asymptomatic CRC screening sample with a sensitivity of 82% and specificity of 85%. In the field of liquid biopsy, AI impact has been examined with the use of supervised learning methods such as support vector machines (SVMs). Hierarchical classification [37] has shown feasible applicability, with an initial classification level filtering out non-CRC samples and a subsequent level differentiating between non-malignant lesions and a "no findings" class. Additionally, blood biomarkers have been considered alongside risk factors (e.g., demographics) within an AI-based framework for estimating

Funded by the European Union

cancer risks. These biomarkers were not limited to complex resource-demanding analytics but included standard blood test markers as well [20], such as red cell distribution width [38] (RDW) and anemia findings. Further knowledge can be drawn from electronic health records (EHR), thus incorporating screening in primary healthcare settings [39]. Notably, combining such models with gFOBT seems to contribute to a more than 2-fold increase in cancer detection capability for retrospective data cohorts. Based on this information pool, decision trees can be constructed for producing risk stratification scores and proposing clinical pathways. Finally, about the CUP case mentioned, AI has been used with known metastatic CRC cases as training data, attempting subsequent testing and validation on both metastatic CRC and CUP data [40]. AI-powered virtual assistants can also provide personalised healthcare services and improve communication between patients and care providers [7]. By this principle, AI-based mobile applications such as the Colorectal Cancer Awareness Application (ColorApp®) [41] have sought to foster community education and participation in screening programs, achieving a score of 72 in the System Usability Scale Questionnaire for the Assessment of Mobile Apps. However, regional adaptations and performance surveys beyond usability scores are lacking. In general, despite the above AI advances in CRC risk and progression assessment, the medical community is still skeptical and reluctant to trust the outcomes of machine learning. This is mainly due to most neural network approaches' depth and confusing architecture, regarded as "black boxes" [42]. Explainable artificial intelligence (XAI) is gradually becoming a prerequisite for clinicians and policymakers seeking to instill accountability and medical transparency into AI-assisted decisions for launching trustworthy clinical applications [43].

Risk Factor Analysis

Numerous studies have investigated the association of CRC incidence with demographic, behavioral, and environmental risk factors, including age, sex, and lifestyle. For instance, a 25% higher incidence has been documented for males, varying among countries [2], while even race has been highlighted as a noteworthy parameter [14]. Overall, age comprises the main factor assessed by current guidelines, formulating at-risk groups for recommended screening. Guidelines suggest screening after 50 years, with healthy citizens advised to pursue regular testing by age 75. For people aged 76-85 y.o., guidelines are based on overall health and prior history, while there are no strict recommendations beyond this range. However, clinical practice has shown that these thresholds are gradually decreasing, a fact under investigation by the medical community. This status creates a pressing need for alternative methods beyond age-only recommendation onset. Currently, the only such criteria correspond to medical history, family history, or symptom manifestation. Indicatively, a meta-analysis of observational studies found that having at least one affected first-degree relative (parents, siblings, or children) increased the risk of CRC by 2.2-fold, and having at least two affected first-degree relatives increased the risk of CRC by almost 4-fold [15].

Moreover, patients with persistent Inflammatory Bowel Disease (IBD) are twice as likely to acquire CRC. Inflammation causes aberrant growth cytokines to be released, as well as increased blood flow, metabolic free radicals, and other variables that contribute to carcinogenesis [16].

Regarding symptoms, CRC may not cause symptoms right away, thus constituting a particularly threatening cancer that necessitates early, reliable screening. Alarming symptoms include rectal bleeding or changes in bowel habits such as diarrhea, constipation, or stool narrowing. In any other case, an individual is assigned a “higher-than-average” risk status under the presence of one (or more) of the following:

- Personal history or family history of CRC or certain types of polyps
- Personal history of inflammatory bowel disease (ulcerative colitis or Crohn's disease)
- Confirmed or suspected hereditary cancer syndrome (2%–5% of all CRC), such as familial adenomatous polyposis coli and its variants (1%), Lynch-associated syndromes (hereditary non-polyposis colon cancer) (2%– 4%), Turcot, Peutz–Jeghers and MUTYH-associated polyposis syndrome
- Personal history of radiation treatment on the abdomen or the pelvic area for a prior cancer

The above elements comprise unalterable facts regarding early detection or even cancer prevention. On the other hand, several lifestyle-related factors have been identified, which are modifiable through suitable behavioral screening and personalised interventions. The links of diet, weight, and exercise to colorectal cancer risk are some of the strongest among all cancer types. For example, being overweight raises the incidence and mortality risk for CRC in both men and women, but the link seems stronger in men. Body mass index (BMI) and waist circumference (WC) are well-established risk factors for CRC, as evidenced by epidemiological research employing a variety of anthropometric measurements [17]. By extension, physical activity seems to constitute a key factor with evidence not favoring a sedentary lifestyle, as is the case for dietary habits. Namely, a diet high in red meats (beef, pork, lamb, or liver) and proccssed meats (like hot dogs and some luncheon meats) is assumed to raise CRC incidence risk. Even cooking-related processes seem to play a part, with very high temperatures (frying, boiling, or grilling) generating chemicals that might raise the associated risk. Similarly, lifestyle habits like smoking or alcohol consumption are linked to CRC incidence [18]. Although smoking is a well-known factor for lung cancer, research has displayed an association with additional malignancies. Similarly, heavy alcohol use may cause several significant health-related outcomes, with colorectal cancer bearing a connection. By extension, all these modifiable factors have the potential to be addressed via lifestyle interventions promoting healthy behaviors, including physical activity, BMI control, appropriate eating/cooking habits, and refraining from smoking or excessive alcohol. Although a large pool of risk factors has been assumed to correlate with colorectal cancer (among other malignancies), the underlying regulatory processes remain largely unknown. The key to quantifying the corresponding transition mechanisms might lie in specific biomarkers that are extracted minimally invasive and cost-effectively. Indicatively, it is a well-known fact that dietary habits are translated into alterations in routine blood biomarkers, such as the blood level of vitamin D, directly associated with eating patterns [19]. However, such phenotype manifestations have not been analysed within a CRC-centered protocol. Some studies have assessed some indicative biomarkers [20] without proposing an interpretation model for behavior effects on biomarker alteration and CRC incidence.

2.2 RATIONALE OF THE STUDY

DIOPTRA aims for an accessible and less taxing screening to attain a broader population outreach by exploiting blood-based biomarkers. Although several researchers have tried to assess this, the limitations in the number of participants and the 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 eight

Funded by the European Union

different sites, utilising many patients/ healthy citizens. As such, the evidence produced will be based on expert opinion and 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 element 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).

3 RETROSPECTIVE CLINICAL STUDY PROTOCOL

This section presents the retrospective study protocol and the bioethics permissions obtained by the clinical partners in their countries.

Clinical Study Protocol	
Clinical protocol number	1
Document version	1.0
Clinical investigation Title	Retrospective data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms (Colorectal Clinical Study)
Novel solution	New technology – DIOPTRA – for cancer screening and early detection
Organisation responsible for clinical investigation (S)	Hospital Name, address
Local representative, if applicable	Name, address
Lead Principal Investigator (I)	Name, address
Coordinating investigator (CI)	Name, address

The following persons are planned to be involved in the clinical investigation:

Role	Organisation responsible for clinical investigation (S) / Principal Investigator (I)/ Coordinating Investigator (CI)	Name	Title
Author			
Clinical Leadership			
Clinical investigation team			
Clinical investigation team			
Clinical investigation team			
Clinical investigation team			
Clinical investigation team			
Clinical Quality			
Regulatory affair			
Medical writer			
Clinical investigation Manager			

Approval sheet

_____ _____

(position) (name, surname)

_____ _____

(signature) (date)

DOCUMENT HISTORY

Revision	Date of enactment	Change author	Change description
1.0			Document is created

3.1 SYNOPSIS

Title	Retrospective data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms (Colorectal Clinical Study)
Purpose	The purpose of this study aims to generate a homogenised dataset consisting of retrospective data on demographic, medical and family history information derived from the Electronic Health Records of the clinical sites for the risk factor investigation for predicting early Colorectal Cancer (CRC) risk based on the risk factor analysis.
Retrospective study design (RSD)	Structured and anonymised retrospective data collected from eight clinical sites.
RSD Primary objective	To analyse all the risk factors and to determine the best features as input to the AI models for predicting early CRC risk.
RSD Secondary objective	To collect, formulate, curate, and harmonise the retrospective data of DIOPTRA.
RSD Primary endpoint	Reveal a panel of risk factors predicting early CRC risk based on statistically significant differences derived from a statistical analysis method.
RSD Secondary endpoint	1) Machine/Deep Learning algorithms based on the accuracy, precision, sensitivity, and specificity of the models 2) Homogenised retrospective DIOPTRA dataset 3) Data curation, data management solution, and infrastructure
Duration of clinical investigation	34 months
Subject population	In total, more than 5000 subjects are estimated to be included across all the study's 8 clinical sites: 1. BLOCKS 2. CHUL 3. RM-RRH 4. UKCM 5. BURGOS 6. NKUA 7. GOC 8. AGSAVVAS

	Gender distribution in the incidence of CRC will be taken into account in this study, including as similar rates of male and female participants as possible. However, it must be taken into account that males are 25% more prone to develop CRC in comparison to females, resulting in more male study participants. The following subject groups will be included in this retrospective study: 1. Healthy: no findings after a colonoscopy or presence of hyperplastic polyps 2. Non-advanced adenomas 3. Advanced adenomas. Under ESGE 2020 guidelines, the following adenoma should be classified as advanced adenomas: at least 1 adenoma ≥ mm or with high-grade dysplasia, or ≥ 5 adenomas, or any serrated polyp ≥ 10 mm or with dysplasia 4. Colorectal cancer CRC stage I, II, and III Colorectal cancer diagnosis with or without surgery, without neo-adjuvant chemotherapy and/or radiotherapy treatment.
Number of subjects	In total, more than 5000 subjects are necessary to be included from 8 clinical sites to account for the minimum sample size per group (sample size justification) and the low incidence rate of CRC (<10%) (BLOCKS, CHUL, RM-RRH, UKCM, BURGOS, NKUA, GOC, AGSAVVAS).
Number of Sites	8 clinical sites: 1) BLOCKS 2) CHUL 3) RM-RRH 4) UKCM 5) BURGOS 6) NKUA 7) GOC 8) AGSAVVAS
External organisations involved in the clinical investigation	No external organisations will be involved in the clinical investigation.
Inclusion criteria	▪ Any indication for total colonoscopy (including routine screening and presence of symptoms/ FIT positive) ▪ Age between 18-80 years at the moment of colonoscopy ▪ Absence of significant comorbidity ASA IV
Exclusion criteria	▪ Age under 18 y/o or above 80 y/o ▪ Comorbidities ASA IV ▪ Recent major abdominal surgery (including colectomy) or radiation prior to the colonoscopy ▪ Inflammatory bowel diseases ▪ Polyposis syndrome

Funded by the European Union

	▪ Pregnancy or suspicion of pregnancy ▪ Colorectal cancer history
Data Oversight	Seeking to ensure the protection of the personal data of the participants in a clinical investigation in accordance with the General Data Protection Regulation (GDPR), it is planned to encrypt personal data: the name of a patient included in the study group will be replaced by a code using a secret key only accessible by the research team in each hospital. Thanks to this secret key, the research team will be the only one to be able to re-identify the patient. Subjects' data collected during the study are included in their personal medical records and stored in the same way as other personal medical data and information. The investigators will transfer the structured and encrypted data required for the study by applying anonymisation tools to the members of the DIOPTRA consortium who will use them. Documents collected for the purpose of this clinical investigation will be kept by the consortium until the end of the project DIOPTRA.
Clinical investigation financing	This Clinical study is part of the DIOPTRA European Project funded within the research and innovation program of the Horizon Europe under N° 101096649
Person paying compensation for costs and time incurred in participating in a clinical investigation, procedure and conditions for calculation and payment of compensation	No compensation is provided.
Abbreviations and acronyms	AGSAVVAS - Geniko Antikarkiniko Ogkologiko Nosokomeio Athinon O Agios Savvas BLOCKS - Blocks Health and Social Care EOOD BURGOS - Fundacion Burgos Por La Investigacion De La Salud CHUL - Centre Hospitalier Universitaire De Liege CIP – clinical investigation protocol CRC - Colorectal Cancer GOC - Linac-Pet Scan Opco Limited NKUA - Ethniko Kai Kapodistriako Panepistimio Athinon RM-RRH - Region Midtjylland UKCM - Univerzitetni Klinicni Center Maribor

3.2 GENERAL INFORMATION

3.2.1 Rationale for the Retrospective Study

Recent advances have contributed to the development of cutting-edge molecular biology tools. As such, epigenetic alterations have been found to initiate and sustain specific deregulated cellular signalling pathways involved in CRC tumours [1]. In this regard, the field of CRC biomarker extraction

from blood samples for screening purposes has been enriched with numerous studies demonstrating a high TRL justifying DIOPTRA ambition for the formulation of a discriminative protein subset.

DIOPTRA aims for an accessible and less taxing screening to attain a wider 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 in 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 9 different sites, utilising a large number of patients/ healthy citizens. As such, the evidence produced will not only be based on expert opinion but from vigorous validation procedures on the retrospective and prospective data (level of evidence B). More importantly, by employing cutting-edge in-vitro protein analysis in (paired with the blood samples collected) biopsies, the biomolecular 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).

Hospitals in 7 countries (BLOCKS, CHUL, RM-RRH, UKCM, BURGOS, NKUA, GOC, AGSAVVAS) will be used for subjects personal medical data inclusion. All clinical sites have a vast amount of data storage and access to a high number of possible participants.

3.2.2 Background

Incidence & Survival Rates. Colorectal cancer [3], [4] (CRC) is the third most common tumor in men and the second in women, accounting for 10% of all tumors 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 from 30.9% to 60% in women, with pooled estimations 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 [5].

Existing Standard & Screening Impact. In the CRC treatment domain, options include endoscopic and surgical excision, radiotherapy, immunotherapy, palliative chemotherapy, targeted therapy, extensive surgery, and local ablative therapies for metastases [3]. Meanwhile, screening methods consisting of endoscopic tests (e.g., colonoscopy) and non-invasive alternatives such as the fecal immunochemical test (FIT) have been put into action.

Pressing Conditions. In determining the CRC risk status, factors such as age, BMI, diet, smoking habits, and family history [6] 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. In this vein, liquid biopsy appears as a promising new tool for non-invasive, quick, and safe assessment [7]. Among all liquid biopsy products, blood-derived proteins seem to constitute the most cost-effective solution judging by resources, sensitivity, and

Funded by the European Union

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 [8]. Generally, outcome prediction has been the hallmark regarding AI utilisation on cancer. Indicatively, AI models have attempted to predict 5-year CRC recurrence risk, outperforming grading and/or staging evaluation by expert pathologists. Relying on the duration of the gradual transition path from normal mucosa to a premalignant growth and then to a malignant lesion spanning over 15-20 years, AI can detect suspect changes corresponding to abnormal tissue, which may be indicative of either a premalignant precursor lesion or an early-stage tumor. Despite the AI advances in CRC risk and progression assessment, the medical community is still skeptical and reluctant to trust the outcomes of machine learning. This is mainly due to the depth and confusing architecture of most neural network approaches, which are regarded as “black boxes” [9]. Explainable artificial intelligence (XAI) is gradually becoming a prerequisite for clinicians and policymakers seeking to instill accountability and medical transparency into AI-assisted decisions for launching trustworthy clinical applications [10].

Risk Factor Analysis. Numerous studies have investigated the association of CRC incidence with demographic, behavioral, and environmental risk factors including age, sex, and lifestyle. Overall, age comprises the main factor assessed by current guidelines, formulating at-risk groups for recommended screening. Clinical practice has shown that these thresholds are gradually decreasing, a fact under investigation by the medical community. Several lifestyle-related factors have been identified, which are modifiable through suitable behavioral screening and personalised interventions. Although a large pool of risk factors has been assumed to correlate with colorectal cancer (among other malignancies as well), the underlying regulatory processes remain largely unknown. The key to quantifying the corresponding transition mechanisms might lie in specific biomarkers that are extracted in a minimally invasive and cost-effective manner.

3.3 STUDY PLAN

3.3.1 Study Hypothesis

The main study hypothesis is that there exists a set of predictive variables that are associated with the risk of developing CRC. Based on this study, risk factors will be identified to examine their association with CRC and predict early CRC risk. Data from the Electronic Health Records will be utilised and analysed to extract variables defined as risk factors based on the four groups. Various methodologies, including statistical analysis and machine learning techniques, will be employed to investigate the influence of each factor on CRC.

3.3.2 Study Design

SAMPLE SIZE JUSTIFICATION

The different risk factors (categorical and numerical) will be tested in terms of their association with the DIOPTRA study groups. In terms of categorical variables, using the chi-squared test, the minimum sample size required per group was calculated using the following assumptions:

- Required power of 0.8
- Significance level: 0.05
- Effect size: 0.1
- Two-sided differences
- Degrees of freedom: 3
- Safety factor: 1.2 (to account for categorical variables with a higher degree of freedom)
- Dropout rate: 25% (to account for the exclusion of participants' records due to errors or missing key variables

The minimum required sample size to detect differences in categorical risk factors is N = 410 samples per group (for the whole study). Additionally, the proposed sample size was used to calculate the power of ANOVA as the base statistical test to identify significant differences in numerical risk factors between study groups. The proposed minimum sample size per group (N=410) results in 0.94 statistical power for the ANOVA test (similar assumptions), which is adequate to detect differences between the groups. Finally, the proposed minimum sample size (N=410, with similar assumptions) results in a precision of less than 5% for the calculation of sensitivity and specificity of the AI risk factor model, assuming both metrics are higher than 60%. All sample size calculations and power analysis were performed using the package pwr in the R statistical language.

DESCRIPTION AND TIMING OF STUDY PROCEDURES

An overview of the study schedule is provided in the table below, followed by all clinical sites.

Table 1: Overview of the study schedule.

No	Description	Timing
1	Variable list definition	M0
2	Data size, availability and quality estimation	M0
3	Data extraction and transformation into structured data	M1-M6
4	Data anonymization and transfer (End of clinical participation)	M7
5	Data processing and analysis by technical partners	M8-M34

STUDY PROCEDURES:

1. Variable list definition

General clinical preparation and CRC expert meetings have already been held in order to elucidate a more suitable variable list for DIOPTRA. The inclusion and exclusion criteria have been modified to be applicable to the study too. Demographic data, medical and family history, clinical biology, symptoms, and detailed diagnosis of CRC are included in the final list. Data templates and corresponding encoding instructions are provided to all clinical partners.

2. Data size, availability, and quality estimation

According to DIOPTRA inclusion and exclusion criteria, relevant participants' electronic health data are recruited for the retrospective study. The sample size, variable availability, and data quality are reported, and the exact recruited data size will be coordinated within the consortium to have a better data balance per group for DIOPTRA.

3. **Data extraction and transformation into structured data**

Data will be extracted, and unstructured data will be transformed into structured data by keyword searching, manual encoding, etc.

4. **Data anonymisation and transfer**

To ensure the privacy of the data, technical partners will provide anonymisation tools to the clinical partners in order to anonymise the retrospective data. These tools will be utilised by clinical partners. The main objective of the current anonymisation tool is to apply the k-anonymity method using the Mondrian algorithm on the input data provided by the clinical partners. Regarding the transferring of the data, an interface will be developed where heterogeneous clinical sites will upload their data in a standardised format. After the harmonisation process, the data will be stored inside an external infrastructure, providing API calls retrieval and additional services.

5. **Data analysis**

When the data is uploaded to the clinical interface, curation, and harmonisation techniques will be employed to fix missing values, identify problematic data fields, and provide a common data model, thereby enhancing the quality and integrity of the data. The outcome will be a curated and harmonised dataset, which will be used for the risk factor analysis. Risk factors will be identified to examine their association with CRC and predict early CRC risk. Advanced Machine Learning algorithms along with statistical analysis will be employed to examine the importance of each factor and rank the most important factors that affect the CRC. Visualisation plots will be developed to increase the transparency and interpretability of machine learning models.

STATISTICAL DATA ANALYSIS

The DIOPTRA methodological approach has been separated into distinct implementation phases. The first phase involves the retrospective collection of tissue/liquid biopsies, demographic, behavioral, environmental, medical, and family history data. The clinical partners will provide Electronic Health Records (EHR) in a structured format based on a reference model, which includes terminologies that are linked with SNOMED CT codes. For the anonymisation process, technical partners will provide specific tools to clinical sites in order to ensure the anonymisation of data prior to transfer. During this phase, curation techniques will be employed to address missing values and identify problematic data fields, improving the accuracy and integrity of the data. Both conventional methods, such as the z-score, the interquartile range, and the Grubb's test, and machine learning-based outlier detection methods, such as the isolation forests and the Gaussian elliptic envelopes, will be utilised to track down and remove values that deviate from the standard distribution on a feature basis. The Logstash that will be utilised is an open server-side data processing pipeline that will deal with noisy and incomplete data.

This phase will result in the DIOPTRA homogenised dataset, which will be utilised in the second phase, where the risk factor analysis and the initial stage of liquid biopsy biomarker analysis will be conducted. On the basis of this dataset, risk factors will be identified to examine their association with CRC. Various methodologies, including statistical analysis (chi-squared test, ANOVA, inverse-variance weighted averages for dichotomous factors, generalised least squares for dose–response for multi-level factors, principal components factor analysis, extraction of behavior patterns) and machine learning techniques (Support Vector Machines, Boosting ensembles, Convolutional Neural Networks, and Long-Short Term Memory algorithm) will be employed to investigate the influence of each factor on CRC. SHapley Additive exPlanations (SHAP) values, along with representative plots,

will be utilised to increase the transparency and interpretability of machine learning models. During this work, a holistic CRC risk factor set and a set of biomarker 'hits' will be produced to be further evaluated in a subsequent analytics step.

The third phase consists of the first stage of the prospective data collection, including risk factor information and pairs of liquid/tissue samples. The extensive pool of liquid/tissue biopsies (consisting of both retrospective and prospective samples) will be analysed during the biomarker development phase to generate protein features for training and evaluating the DIOPTRA AI model. The focus of the fourth phase will be on the latter process, in which risk factor and protein features will be utilised in an AI-based analysis to generate the proposed subset of discriminative protein biomarkers as well as the full stratification model using a hybrid approach based on the combination of biomarker and risk factor phenotypes. In addition, this knowledge will be incorporated into the DIOPTRA mobile application, which will monitor information and suggest individualised behavioral interventions. Data management and processing nodes for AI modules will be set up, utilising the ARIS HPC system with Louros DC as the central node, endorsing EU-based green data centers.

3.3.3 Number of Investigation Sites and Study Duration

The planned duration of the entire clinical investigation: ….months.
Start of the clinical investigation: …….. End of the clinical investigation: ……..
Sites of investigation: 8 sites.

Research center No 1: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …….. ; e-mail: …….. .

Research center No 2: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …….. ; e-mail: …….. .

Research center No 3: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …….. ; e-mail: …….. .

Research center No 4: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …….. ; e-mail: …….. .

Research center No 5: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..

Title of the department (s): ……..
Tel: ……..; Fax: ……. ; e-mail: …….. .

Research center No 6: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: ……. ; e-mail: …….. .

Research center No 7: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: ……. ; e-mail: …….. .

Research center No 8: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: ……. ; e-mail: …….. .

3.4 STUDY METHODS

3.4.1 Data Collection Requirements

Documents collected for the purpose of this clinical investigation will be kept by the Organisation responsible for clinical investigation until the DIOPTRA project will be finished. Data collected during the study will be included in their personal medical records and stored at the Investigator's Office in the same way as other personal medical data and information.

3.4.2 Source Documents

Investigators are required to maintain records of each subject's case history. Source documents include the participant's hospital files (electronic or paper). The investigator will record which subject is enrolled in this clinical investigation.
The investigator(s) and study personnel must be accessible to the clinical support and the clinical study team. This accessibility is of particular importance for the completion and clarification of the regulatory responsibilities. Access to the subject records and other source data must be provided to study monitors, auditors, and/or inspectors.

3.4.3 Study Deviations and Clinical Study Protocol Changes

A study deviation is an event where the investigator or investigation site personnel did not conduct the clinical study according to the Clinical Study Protocol. The investigator is not allowed to deviate from the above-mentioned documents except with prior approval and under emergency circumstances. All deviations shall be documented and explained, regardless of the reason for the deviation.

Funded by the European Union

3.5 QUALITY CONTROL PROCEDURES

3.5.1 Data review and processing

During the study, the completeness of patient records will be checked based on the accuracy of entries, the adherence to the protocol and to Good Clinical Practice as well as GDPR protocols, and the progress of enrolment.
Data management will be done according to "……" internal procedures and the Data Management Plan for this clinical investigation. These documents 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 investigator to complete, correct or comment on the data.

3.5.2 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 the suspension or termination of the protocol procedures.

In this case, the investigator(s) must inform the Organisation responsible for clinical investigation of the reasons for the termination of the study, and the data collected prior to the termination of the study must be passed on to the Organisation responsible for clinical investigation.
Any changes will be agreed in advance with the Bioethical Committee that authorised the clinical investigation.

3.5.3 Study Close Out

"……" and/or its designees 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 investigator's regulatory files are up to date and complete and that any outstanding issues from previous visits have been resolved. "……" will notify and inform the site(s) that all requirements have been met with a study closure letter.
"……" will notify Bioethical Committee about the clinical trials closure by providing a Clinical Studies report based on the Bioethical Committee/another regulatory authority form.

Funded by the European Union

3.6 DATA REPORTING AND PUBLICATION

3.6.1 Data Reporting and Publication

Any deviations from the CIP will be described and justified in the Final Clinical Study Report, as appropriate.
Publications and presentations referring to this clinical study will be coordinated by "……." to allow the use of all available data.

3.7 RISKS AND BENEFITS

3.7.1 Risk-to-Benefit Rationale

Risk-to-benefit rationale data are provided in the table 1 below:

Table 1 Risk-to-benefit rationale data

Extent of impact	Benefit	Risk	Conclusion
To the patient	§ Early detection of CRC can significantly enhance survival rates and treatment outcomes § Identifying high-risk individuals allows healthcare providers to offer personalised prevention plans to reduce the risk of developing CRC	§ For the purposes of the project, a vast and varied amount of personal data will be collected, stored, and processed, including special category data (medical data). There is a risk that personal medical data will be lost/accessible to third parties. The risk is very low because personal data management will be ensured in accordance with the legal requirements. § Ethical issues may arise if, throughout the monitoring procedures and the collection of health-related data, unexpected observations of clinical significance are discovered ('incidental findings'). Findings such as these have potentially serious implications for a participant's medical health, psychological well-being, or even employment and may also affect the subject's relatives	The benefits outweigh the risks
To the medical institution	§ Effective prevention strategies and interventions, such as lifestyle modifications, dietary recommendations, and chemoprevention options, would	§ Loss of personal data would damage the reputation of the institution. The risk is very low because personal data management will be ensured in accordance with the legal requirements.	The benefits outweigh the risks

	reduce the number of patients in medical institutions		
To the country medical system	§ Risk factor analysis can provide valuable insights into the modifiable risk factors for CRC. This information can be used to develop effective prevention strategies and interventions, such as lifestyle modifications, dietary recommendations, and chemoprevention options, which would save health system costs for treatment	§ Risk is not detected.	The benefits outweigh the risks

3.8 BIBLIOGRAPHY

[1] 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: Apr. 18, 2023. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcell.2023.660924
[2] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era," Curr. Oncol., vol. 28, no. 3, pp. 1581– 1607, Apr. 2023, doi: 10.3390/curroncol28030149.
[3] Y. Xi and P. Xu, "Global colorectal cancer burden in 2020 and projections to 2040," Transl. Oncol., vol. 14, no. 10, p. 101174, Apr. 2023, doi: 10.1016/j.tranon.2023.101174
[4] 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, Apr. 2023, doi: 10.1016/j.annonc.2023.04.022.
[5] 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. 2021, doi: 10.1002/cam4.1697.
[6] 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, Apr. 2023, doi: 10.1038/s41598-020- 76294-w.
[7] 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, Apr. 2021, doi: 10.3390/ijms19113356.

[8] 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, Apr. 2021, doi: 10.1016/j.csbj.2021.04.001.
[9] 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, Apr. 2023, doi: 10.1109/TNNLS.2023.3027314.
[10] 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, Mar. 2023, doi: 10.1016/j.jbi.2023.103523.

Bioethics Committee permission of CHUL

Comité d'Ethique Hospitalo-Facultaire Universitaire de Liège (707)

Sart Tilman, le 16/05/2023

Monsieur le **Prof. E. LOUIS**
Madame le **Dr C. LOLY**
Service de **GASTROENTEROLOGIE - HEPATOLOGIE - ONCOLOGIE DIGESTIVE**
CHU B35

Concerne: Votre demande d'avis au Comité d'Ethique
Notre réf: 2023/130

"Dépistage dynamique précoce du cancer colorectal (CCR) grâce à de nouveaux biomarqueurs protéiques reflétant les mécanismes d'initiation biologique (DIOPTRA). "
Protocole : **V1**

Cher Collègue,

Le Comité d'Ethique constate que votre étude n'entre pas dans le cadre de la loi du 7 mai 2004 relative aux expérimentations sur la personne humaine.

Le Comité n'émet pas d'objection éthique à la réalisation de cette étude.

Vous trouverez, sous ce pli, la composition du Comité d'Ethique.

Je vous prie d'agréer, Cher Collègue, l'expression de mes sentiments les meilleurs.

Prof. V. SEUTIN
Président du Comité d'Ethique

C.H.U. de LIEGE – Site du Sart Tilman – Avenue de l'Hôpital, 1 – 4000 LIEGE
Président : Professeur V. SEUTIN
Secrétaire exécutif : Docteur G. DAENEN
Secrétariat administratif : 04/323.21.58 – Coordination scientifique: 04/323.22.65
Mail : ethique@chuliege.be
Infos disponibles sur:http://www.chuliege.be/orggen.html#ceh

Funded by the European Union

MEMBRES DU COMITE D'ETHIQUE MEDICALE
HOSPITALO-FACULTAIRE UNIVERSITAIRE DE LIEGE

Monsieur le Professeur **Vincent SEUTIN** — **Président**
Pharmacologue, membre extérieur au CHU

Monsieur le Docteur **Guy DAENEN** — **Secrétaire exécutif**
Honoraire, Gastro-entérologue, membre extérieur au CHU

Monsieur **Resmi AGIRMAN**
Représentant des volontaires sains

Monsieur le Docteur **Etienne BAUDOUX**
Expert en Thérapie Cellulaire, CHU

Madame **Viviane DESSOUROUX** / Monsieur **Pascal GRILLI** (suppléant)
Représentant (e) des patients

Monsieur le Professeur **Pierre FIRKET**
Généraliste, membre extérieur au CHU

Madame **Régine HARDY** / Madame la Professeure **Adélaïde BLAVIER** (suppléante)
Psychologue, CHU — Psychologue, membre extérieure au CHU

Madame **Isabelle HERMANS**
Assistante sociale, CHU

Monsieur le Professeur **Maurice LAMY**
Honoraire, Anesthésiste-Réanimateur, membre extérieur au CHU

Madame la Docteure **Marie-Paule LECART**
Rhumato-gériatre, CHU

Monsieur le Docteur **Didier LEDOUX**
Intensiviste, CHU

Madame **Marie LIEBEN**
Philosophe, membre extérieure au CHU

Madame **Patricia MODANESE**
Infirmière cheffe d'unité, CHU

Madame la Professeure **Anne-Simone PARENT**
Pédiatre, CHU

Monsieur le Professeur **Marc RADERMECKER**
Chirurgien, CHU

Monsieur **Stéphane ROBIDA**
Juriste, membre extérieur au CHU

Madame **Isabelle ROLAND**
Pharmacien, CHU

Madame la Docteure **Liliya ROSTOMYAN**
Endocrinologue, CHU

Madame la Docteure **Isabelle RUTTEN**
Radiothérapeute, membre extérieure CHU

Madame **Cécile THIRION**
Infirmière cheffe d'unité, CHU

Bioethics Committee permission of BURGOS

Complejo Asistencial Universitario de Burgos
Avda. Islas Baleares, 3 - 09006 BURGOS
Teléfono 947 28 18 00

Sacyl

Informe del Comité de Ética de la Investigación con Medicamentos

Don Jorge Labrador Gómez Secretario Técnico del Comité de Ética de Investigación con Medicamentos del Área de Salud Burgos y Soria,

CERTIFICA:

Que este Comité ha informado de la propuesta para que se realice el Estudio, titulado: **"Recopilación retrosprectiva de datos para el cribado dinamico precoz del cáncer colorrectal mediante nuevos biomarcadores proteinicos que reflejan los mecanismos biológicos de iniciación. Colorectal Clinical Stady. DIOPTRA"** (*Ref. CEIm 2960)* para que sea realizado por el Dr. Enrique Lastra Aras del Servicio de Oncología Médica del Hospital Universitario de Burgos como investigador principal.

Este comité constata qué a dicho Estudio, no le es de aplicación el Real Decreto 1090/2015 de Ensayos Clínicos con Medicamentos. Este CEIm de Área de Salud de Burgos y Soria se da por enterado.

Lo que firmo en Burgos, 30 de mayo de 2023

D. Jorge Labrador Gómez

Junta de Castilla y León

Bioethics Committee permission of GOC

ΚΥΠΡΙΑΚΗ ΔΗΜΟΚΡΑΤΙΑ

ΕΘΝΙΚΗ ΕΠΙΤΡΟΠΗ ΒΙΟΗΘΙΚΗΣ ΚΥΠΡΟΥ

Αρ. Φακ.: ΕΕΒΚ ΕΠ 2023.01.147
Αρ. Τηλ.: 22809038/039, 22819101/122
Αρ. Φαξ: 22353878

8 Ιουνίου, 2023

Δρ Κρίστης Βέβης
Διαχειριστής Ερευνητικών Προγραμμάτων
Γερμανικό Ογκολογικό Κέντρο
Λεωφ. Νίκης 1
4108 Άγιος Αθανάσιος
Λεμεσός

Αγαπητέ Δρ Βέβη,

Αίτηση γνωμοδότησης για την πρόταση με τίτλο:
«DIOPTRA - Retrospective data collection for early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms»

Αναφορικά με την αίτηση σας ημερομηνίας 2 Ιουνίου 2023 για το πιο πάνω θέμα, επιθυμώ να σας πληροφορήσω ότι από τη μελέτη του περιεχομένου των εγγράφων που έχετε καταθέσει η Εθνική Επιτροπή Βιοηθικής Κύπρου (ΕΕΒΚ) **γνωμοδοτεί θετικά υπέρ της διεξαγωγής της εν λόγω έρευνας.**

2. Η Επιτροπή επιθυμεί να τονίσει ότι παραμένει ευθύνη δική σας η διεξαγωγή της έρευνας με τρόπο που να τηρούνται οι πρόνοιες του νέου Ευρωπαϊκού Γενικού Κανονισμού Προστασίας Προσωπικών Δεδομένων (2016/679) και του περί της Προστασίας των Φυσικών Προσώπων Έναντι της Επεξεργασίας των Δεδομένων Προσωπικού Χαρακτήρα και της Ελεύθερης Κυκλοφορίας των Δεδομένων αυτών Νόμος του 2018 (Ν. 125(Ι)/2018), ως αυτός εκάστοτε τροποποιείται.

3. Σας ενημερώνουμε ότι για σκοπούς καλύτερου συντονισμού και αποφυγής επανάληψης ερευνών με το ίδιο θέμα ή/και υπό εξέταση πληθυσμό μέσα σε σύντομο σχετικά χρονικό διάστημα, η ΕΕΒΚ δημοσιεύει στην ιστοσελίδα της το θέμα της έρευνας, τον φορέα και τον υπό εξέταση πληθυσμό.

4. Κατά τη διάρκεια εκπόνησης της έρευνας, ο συντονιστής / επιστημονικός υπεύθυνος θα ενημερώνει την ΕΕΒΚ για κάθε τροποποίηση των αρχικά κατατεθειμένων εγγράφων (πρωτόκολλο ή άλλα ερευνητικά έγγραφα) και θα υποβάλλει τις απαιτούμενες έντυπες τροποποιήσεις στην Επιτροπή.

5. Σε περίπτωση διακοπής της έρευνας, ο συντονιστής / επιστημονικός υπεύθυνος θα ενημερώσει γραπτώς την Επιτροπή κάνοντας αναφορά και στους λόγους διακοπής της έρευνας.

.../2

Λαέρτου 22, 2365 Άγιος Δομέτιος, Λευκωσία
Ηλεκτρονικό Ταχυδρομείο: enbc@bioethics.gov.cy, ιστοσελίδα: www.bioethics.gov.cy

Funded by the European Union

6. Ο συντονιστής / επιστημονικός υπεύθυνος θα ενημερώσει την Επιτροπή σε περίπτωση αδυναμίας να συνεχίσει ως συντονιστής και θα υποβάλει τα στοιχεία επικοινωνίας του αντικαταστάτη του.

7. Με το πέρας της ερευνητικής πρότασης, ο συντονιστής / επιστημονικός υπεύθυνος θα ενημερώσει εγγράφως την Επιτροπή ότι το υπό αναφορά ερευνητικό πρωτόκολλο ολοκληρώθηκε.

8. Σας ευχόμαστε κάθε επιτυχία στη διεξαγωγή της έρευνάς σας.

Με εκτίμηση,

Καθ. Κωνσταντίνος Ν. Φελλάς
Πρόεδρος
Εθνικής Επιτροπής Βιοηθικής Κύπρου

4 BIOMARKER DISCOVERY STUDY PROTOCOL

Study protocol

Early dynamic screening for colorectal cancer via novel protein biomarkers reflecting biological initiation mechanisms (DIOPTRA)

-

Retrospective biomarker discovery pilot study

Funded by the European Union

4.1 MEDICAL UNIVERSITY OF GRAZ PROJECT PARTNERS

GERGER, Armin, **Assoz. Prof. Priv.-Doz. Dr.med.univ.et scient.med. MBA (PI)**

Division of Oncology

8036 Graz, Auenbruggerplatz 15

E-Mail: armin.gerger@medunigraz.at

KUSE-ISINGSCHULTE, Tatjana, **Dr.med.univ.**

Division of Oncology

8036 Graz, Auenbruggerplatz 15

E-Mail: tatjana.kuse-isingschulte@medunigraz.at

GROLLER, Karin, **MPH**

Division of Oncology

8036 Graz, Auenbruggerplatz 15

E-Mail: karin.groller@medunigraz.at

SCHLEMMER, Andrea

Institute for Medical Informatics, Statistics and Documentation

8036 Graz, Auenbruggerplatz 2/9/III

E-Mail: andrea.schlemmer@medunigraz.at

VALJAN, Monika, **MA**

Biobank Graz Director

8010 Graz, Neue Stiftingtalstraße 2

E-Mail: monika.valjan@medunigraz.at

4.2 TITLE, ACRONYM, UNIQUE IDENTIFIER

Title: early **D**ynam**I**c screening for c**O**lorectal cancer via novel **P**ro**T**ein bioma**R**kers reflecting biologic**A**l initiation mechanisms (Colorectal Clinical Study)

Acronym: DIOPTRA

4.3 COOPERATION PARTNERS

The DIOPTRA project is funded by the European Commission (HORIZON-MISS-2021-CANCER-02) and consists of 28 international project partners including academic and industrial facilities as listed in Table 1.

Table 1: List of international cooperation partners

Number	Role	Short name	Legal name	Country
1	COO	DCHE	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	DK
2	BEN	TCR	TECREANDO B.V.	NL
3	BEN	UOI	PANEPISTIMIO IOANNINON	EL
4	BEN	SIA	SMARTSOL SIA	LV
5	BEN	I2G	I2GROW INNOVATION TO GROW SRL	IT
6	BEN	INTRA	NETCOMPANY-INTRASOFT SA	LU
7	BEN	PAO	PROTAVIO - ETAIREIA EREYNAS VIOTECHNOLOGIAS MONOPROSOPI ETAIREIA PERIORISMENIS EYTHINIS	EL
8	BEN	CSCY	CSCY COMPUTER SOLUTIONS CYPRUS LTD	CY
9	BEN	BLOCKS	BLOKS ZDRAVNI I SOTSIALNI GRIZHI EOOD	BG
10	BEN	ARTHUR	ARTHUR'S LEGAL BV	NL
11	BEN	CSI	CSI CENTER FOR SOCIAL INNOVATION LTD	CY
12	BEN	CHUL	CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE	BE
13	BEN	TERAGLOBUS	UAB TERAGLOBUS	LT
14	BEN	VILABS	VILABS (CY) LTD	CY
15	BEN	RM-RRH	REGION MIDTJYLLAND	DK
16	BEN	UKCM	Univerzitetni klinicni center Maribor	SI
17	BEN	DCHE	KOMITEEN FOR SUNDHEDSOPLYSNING	DK

18	BEN	HOPE	FEDERATION EUROPEENNE DES HOPITAUX ET DES SOINS DE SANTE	BE
19	BEN	NOVELCORE	D.TSAKALIDIS-G.DOMALIS OE	EL
20	BEN	AINIGMA	AINIGMA TECHNOLOGIES	BE
21	BEN	BURGOS	FUNDACION BURGOS POR LA INVESTIGACION DE LA SALUD	ES
22	BEN	NKUA	ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON	EL
23	BEN	GOC	LINAC-PET SCAN OPCO LIMITED	CY
24	BEN	AGSAVVAS	GENIKO ANTIKARKINIKO OGKOLOGIKO NOSOKOMEIO ATHINON O AGIOS SAVVAS	EL
25	BEN	GRAZ	MEDIZINISCHE UNIVERSITAT GRAZ	AT
26	AP	MARTEL GMBH	MARTEL GMBH	CH
27	AP	STS	SPHYNX TECHNOLOGY SOLUTIONS AG	CH
28	AP	CMA	CAMBRIDGE MEDICAL ACADEMY LTD	UK

4.4 BACKGROUND

Colorectal cancer (CRC) [1], [2] is the third most common cancer in men and the second in women, accounting for 10% of all cancers 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. CRC also carries a heavy financial burden (estimated at €19.1 billion across Europe in 2015) [3], due to high costs associated with screening, endoscopic surgical procedures, pharmacological treatment, and the need for lifelong frequent colposcopy follow-up examinations, as well as frequent hospitalisations and the high recurrence rate. Although the etiology of CRC is attributed to multiple risk factors (age, BMI, diet, smoking habits and family history [4]), CRC preventability based on modifiable risk factors, awareness and knowledge exploitation remain extremely low. On another hand, another way to tackle the issue is by increasing population screening and its related measures, which have already proven to be highly effective. Specifically,

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]. Particularly, 5-year survival rate can reach 90% for stage I diagnosis, being less than 15% for advanced stages [6]. Therefrom, routine screening is key for reducing mortality and declining incidence rates, since CRC is now considered as a highly preventable disease with a considerably wide temporal development window [7]. Therefrom, routine screening is key with the current CRC standard of care consisting of endoscopic tests (e.g. colonoscopy) and non-invasive alternatives (faecal immunochemical test (FIT)). However, taxing procedures, citizen reluctance, poor awareness and screening accessibility are hindering participation [8], 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.

DIOPTRA aspires to incorporate CRC screening services into standard bloodwork, broadening the evaluated population, while providing the foundations of CRC development risk factor assessment, thus boosting participation and bypassing age screening thresholds. In this study, we intend to identify specific CRC protein biomarkers derived from blood samples of patients and (health controls) citizens, utilising Artificial Intelligence (AI) tools and clinically evaluating them via endoscopic colonoscopy. Moreover, employing cutting-edge biomolecular procedures on paired CRC tissue, we aim to uncover the CRC development underlying mechanism.

4.5 OBJECTIVES

The objective of the DIOPTRA discovery study is to discover a panel of blood-based diagnostic protein biomarkers, with verified connection to the colorectal cancer mechanism, for CRC screening and early detection.

The primary and secondary outcomes of the prospective DIOPTRA Clinical studies at the other sites are:

(i) *Primary outcome*: Delivery of an efficient CRC screening protocol, via thorough investigation of minimally-invasive liquid biopsy, along with strong (currently exceedingly limited) evidence of the underlying mechanism regulating CRC.

(ii) *Secondary outcome #1* – Collect a comprehensive CRC dataset from patients.

(iii) *Secondary outcome #2* – Improve overall quality of life of CRC patients.

(iv) *Secondary outcome #3* – Reduce unnecessary invasive screening procedures.

(v) *Secondary outcome #4* – Increase CRC awareness and screening programs participation.

4.6 HYPOTHESIS

We hypothesise that a distinct set of protein biomarkers is present in blood, which is associated with the underlying mechanisms of colorectal cancer (CRC), and accurately differentiates between individuals in the four study groups: healthy, non-advanced adenoma, advanced adenoma, and colorectal cancer. This difference will be detectable through comprehensive analysis of blood and tissue samples using proteomics, transcriptomics, and computational analysis methods.

- **Null hypothesis H_0:** There is no statistically significant difference in the serum biomarker profiles between healthy individuals and those with early-stage CRC.
- **Alternative hypothesis H_1:** The serum biomarker profiles between healthy individuals and CRC patients are significantly different, and can be utilised for cancer presence assessment and prediction.

4.7 SAMPLE SELECTION

Retrospective data and samples from up to 200 subjects will be acquired, corresponding to four study groups (50 subjects each), including:

(i) Healthy control citizens: no findings after a colonoscopy or presence of hyperplastic polyps

(ii) Non-advanced adenomas

(iii) Advanced adenomas

(iv) Colorectal cancer (CRC) stage I, II III and IV

Colorectal cancer diagnosis with or without surgery, without neo-adjuvant chemotherapy and/or radiotherapy treatment.

Biological samples and data will be acquired in a group-specific manner as shown in Table 2. Tissue and liquid samples will be provided paired if possible.

Gender distribution in the incidence of CRC will be taken into account in this study, aiming for similar rates of male and female participants. However, it must be taken into account that males are 25% more prone to develop CRC in comparison to females, likely resulting in more male study participants. It is also planned to include retrospective data from subjects without corresponding

Funded by the European Union

samples. If a final plan is established, the number of subjects and the variables to be collected will be specified in an amendment.

Table 2: Sample and data acquisition scheme

	Healthy	**Non-advanced adenomas**	**Advanced adenomas**	**CRC**
Tissue biopsy (cryo-frozen)	*Not applicable*	Yes	Yes	Yes
Liquid biopsy (serum)	Yes	Yes	Yes	Yes
Corresponding data of same subject	Yes	Yes	Yes	Yes

4.8 INCLUSION CRITERIA

1. Any indication for total colonoscopy (including routine screening and presence of symptoms/ FIT positive)
2. Age between 18-80 years at the moment of colonoscopy
3. Absence of comorbidity ASA IV

4.9 EXCLUSION CRITERIA

1. Age under 18 y/o or above 80 y/o
2. Comorbidities ASA IV
3. Radiation prior to the tissue sample
4. Inflammatory bowel diseases
5. Polyposis syndrome
6. Pregnancy or suspicious of pregnancy
7. Colorectal cancer history

4.10 STUDY DESIGN: STUDY TYPE

The study is going to be a retrospective biomarker discovery pilot study. Already collected samples and data from participants of all four groups will be used. Initial data obtained will be used for the algorithm training followed by validation of the pilot. This biomarker discover study is followed by a prospective study, which will be conducted at the other clinical sites of this EU project.

Funded by the European Union

4.11 STUDY DESIGN: DATA COLLECTION

The Institute for Medical Informatics will perform a pre-selection of potential study participants based on their CRC history, available samples and consent to Biobank IC. Eligible ICD_10 codes for the diseased cohort are C18.0, C18.1, C18.2, C18.3, C18.4, C18.5, C18.6, C18.7, C19.9, C20, C20.9, C21, C21.0, C21.1, C21.2, C18.2, C18.6, C18.9, C18.8, C21.8. As some study-relevant data cannot be accessed this way, the final patient selection will be done by a trained employee of the Clinical Division for Oncology at the Medical University of Graz by using the patient identifier for a detailed open MEDOCS request. For the other groups, existing biobank cohorts will be screened for participants who had a coloscopy in the clinical setting according to in- and exclusion criteria. Clinical information of the human subject will be used to categorise the biological specimens into the four categories of the DIOPTRA study (healthy, non-advanced adenoma, advanced adenoma and CRC). This characterisation will be performed by the Division of Oncology in collaboration with the Institute for Medical Informatics, Statistics and Documentation. Subsequently, the data will be provided to the test facility "PROTAVIO Ltd." and to the technical coordinator "PANEPISTIMIO IOANNINON". Clinical characterisation is based on the clinical data as described in Table 3**:** Clinical data collected for the DIOPTRA study.

Table 3: Clinical data collected for the DIOPTRA study

Basic patient information
Patient ID, Age in years, Sex, Date of blood drawing, ICD10, Age at primary diagnosis in years, Diagnosis date, Sample type, Height, Weight, BMI, Comorbidities, Medication,
Oncological data
Colonoscopy (incl. date), Metastasis (y/n), Localisation of primary tumor, Morphology, UICC Version, UICC Stage, TNM Primary tumor, TNM Regional Lymph nodes, TNM Distant Metastasis, WHO Grade, Karnofsky Index, ECOG Status, Histological diagnosis
Molecular markers
Microsatelite Status, BRAF, KRAS, NRAS

4.12 STUDY DESIGN: MATERIAL COLLECTION

Serum and fresh frozen cryo tissues have already been collected in the past and stored at Biobank Graz at -80°C (serum) and in the gas phase of liquid nitrogen at -150 to - 160°C (cryo tissue).

Sample transport

Before sample shipping, a "Material Transfer Agreement" between sender (Med Uni Graz) and recipient (PROTAVIO Ltd.) will be signed. All biological samples will be shipped to PROTAVIO in DRY ICE using next-day delivery Courier services following the information below:

- The description of samples that will be provided to the courier service should be:

  UN3373 Biological Substance Cat B packed in Dry Ice, Class 9, UN1845 ….. kgs

  Use for research purposes only

- The shipment will be arranged between Monday and Wednesday to ensure delivery by end of week and sent to the following address:

  PROTAVIO Ltd.
  NCSR Demokritos
  Patriarchou Grigoriou E' & 27 Neapoleos Street,
  Lefkippos Technology Park, Bldg 27
  15341, Agia Paraskevi, Attiki
  Greece
  Tel: +30 2109610307

- On the day of the shipment, care is taken to ensure that all biological samples are covered with dry ice and do not thaw in the process.

Trained personnel at the Test Facility (Protavio Ltd.) will receive biological specimens shipped from Biobank Graz and the appropriate biological specimen receipt form will be completed, dated, and signed. Sample quantity, quality, shipping conditions, unique identifiers and records will be assessed.

Observations including deviations will be recorded accordingly.

4.13 STUDY DESIGN: METHODS AND DATA ANALYSIS

Serum samples from individuals belonging to the four study groups will be analysed with the Olink Explore platform. The Explore platform is based on the Proximity Extension Assay (PEA) that uses matched pairs of antibodies attached to unique DNA nucleotides for each protein allowing hybridisation only to each other. Subsequent proximity extension creates unique DNA reporter

sequences which are amplified by real-time PCR and sequenced using NGS methods. This state-of-the-art proteomics method will allow for the parallel screening of 3000 proteins in minimal volumes of serum samples and their relative quantification in the four study groups. The multiplex proteomics readouts will be pre-processed, quality controlled and normalised using Olink's algorithms in the R statistical language. The processed data will be analysed with statistical and machine learning methods, including PCA, t-SNE, LDA, clustering algorithms, two-sided statistical tests and supervised classification algorithms. The goal of the proteomics analysis is to identify a set of protein biomarkers that show a significant difference in expression between the different study groups (healthy, nonadvanced adenoma, advanced adenoma, colorectal cancer).

From **tissue samples**, RNA and/or protein contents will be extracted and analysed with next generation sequencing methods and/or Olink proteomics. This aims to detect differences between healthy tissue and tissue constituting the primary site of cancer development, in regard to gene expression and/or proteome. A systems biology-based pipeline will be used to combine the gene expression and protein data along with prior knowledge of molecular interactions, from open-source knowledgebases (e.g. Omnipath), to create a protein-protein interaction (PPI) network representation of the CRC mechanism. The resulting PPI network will be analysed using network analysis, including clustering, shortest path and community detection algorithms to evaluate the importance of each node in the network. At this stage, the selected set of blood-based protein biomarkers, discovered via the Olink analysis of matched blood samples, will be mapped to the network and their relation to the CRC mechanism will be verified. The goal of the systems biology pipeline is to select the optimal bloodbased protein biomarkers that show a strong relation to the gene and protein expression patterns at the primary site of cancer development.

The selected panel of protein biomarkers will be developed into highly specific and cost-effective multiplex assays to provide powerful features for the DIOPTRA detection system. For assay development we will utilise the xMAP platform (Luminex Corp). The xMAP technology relies on color-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 relative and absolute quantification of multiple proteins across the various conditions tested. These assays offer high multiplexability, sample throughput, quality of measurements and specificity for measurement of identified biomarkers in blood. During the development process different antibody pairs will be evaluated for each target and the analytical

Funded by the European Union

performance of the assays will be verified using the retrospective serum samples. The multiplex readouts from the developed assays will then be combined with logic-based and AI-based computational models to select the optimal combination of biomarkers that maximises their detection and diagnostic performance.

Retrospective data: The DIOPTRA methodological approach has been separated into distinct implementation phases. The first phase involves the retrospective collection of tissue/liquid biopsies, demographic, behavioural, environmental, medical, and family history data. The clinical partners will provide defined clinical data in a structured format, based on a reference model, which includes terminologies that are linked with SNOMED CT codes. For the anonymisation process, technical partners will provide specific tools to clinical sites in order to assure the anonymisation of data prior to transfer.

The second phase will result in the DIOPTRA homogenised dataset, where the risk factor analysis and the initial stage of liquid biopsy biomarker analysis will be conducted. On the basis of this dataset, risk factors will be identified to examine their association with CRC.

The third phase consists of the first stage of the prospective data collection at other clinical sites, including risk factor information and blood samples. The extensive pool of liquid/tissue biopsies (consisting of both retrospective and prospective samples) will be analysed during the biomarker development phase to generate protein features for training and evaluating the DIOPTRA AI model.

The focus of the fourth phase will be on the latter process, in which risk factor and protein features will be utilised to an AI-based analysis to generate the proposed subset of discriminative protein biomarkers as well as the full stratification model using a hybrid approach based on the combination of biomarker and risk factor phenotypes. In addition, this knowledge will be incorporated into the DIOPTRA mobile application, which will monitor information and suggest individualised behavioral interventions.

4.14 STUDY DESIGN: ARCHIVING

All documentation related to the biological specimens will be archived by the test facility. The remaining quantities of serum and tissue specimens will be stored by the test facility at -80°C in ultralow freezers. Biological specimens and related documentation will be archived for a period of two years following completion of the DIOPTRA project.

Funded by the European Union

4.15 STUDY DESIGN: DISPOSAL

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

4.16 STATISTICAL ANALYSIS

Serum and tissue samples

The following section outlines the key steps involved in the statistical analysis of the **RNA sequencing data** for comparing the expression of genes between the different study groups.

1. Data pre-processing: The first step in the statistical analysis of RNA sequencing data is to preprocess the data to ensure that it is of high quality and ready for analysis. This includes steps such as quality control checks, read alignment, and normalisation of the reads using DESeq2's median of ratios or EdgeR's trimmed mean of M values.

2. Differential expression analysis: The second step of the statistical analysis is to identify differentially expressed genes between the groups being compared. On this front statistical tests such as the t-test, ANOVA, general linear models will be utilised. The statistical analysis will consider factors such as batch effects, multiple comparisons, and adjustment for confounding variables.

3. Enrichment analysis: After identifying differentially expressed genes, enrichment analysis will be performed to determine the biological processes and pathways associated with the identified genes. This can be achieved using tools such as Gene Set Enrichment Analysis (GSEA) or statistical pipelines based on the Kolmogorov-Smirnoff test (KS test).

For the statistical analysis of the **Olink proteomics data**, the following steps will be performed:

1. First, the NPX readouts will be analysed in terms of their quality using quality control plots such as histograms and PCA plots. The Olink data will be normalised using inter-plate controls.

2. In terms of differential protein expression, the data will be analysed using general linear models and the F-statistical test. The results will be visualised using volcano plots that indicate the differential expression of each protein.

3. In addition to univariate analysis, multivariate analysis will be performed to identify a panel of proteins that can differentiate between the groups being compared. This can be achieved using techniques such as principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), or random forest analysis.

4. Finally, pathway analysis using the KS test will be performed to determine the biological processes and pathways associated with the identified differentially expressed proteins.

Funded by the European Union

Retrospective data First phase:

During this phase, curation techniques will be employed to address missing values and identify problematic data fields, improving the accuracy and integrity of the data. Both conventional methods, such as the z-score, the interquartile range, and the Grubb's test, and machine learning based outlier detection methods, such as the isolation forests and the Gaussian elliptic envelopes, will be utilised to track down and remove values that deviate from the standard distribution on a feature basis. The Logstash that will be utilised is an open server-side data processing pipeline that will deal with noisy and incomplete data.

Second phase:

Various methodologies, including statistical analysis (inverse-variance weighted averages for dichotomous factors, generalised least squares for dose–response for multi-level factors, principal components factor analysis, extraction of behavior patterns) and machine learning techniques (Support Vector Machines, Boosting ensembles, Convolutional Neural Networks and Long-Short Term Memory algorithm) will be employed to investigate the influence of each factor on CRC. SHapley Additive exPlanations (SHAP) values along with representative plots will be utilised to increase transparency and interpretability of machine learning models. During this work, a holistic CRC risk factor set and a set of biomarker 'hits' will be produced to be further evaluated in a subsequent analytics step.

4.17 ETHICAL CONSIDERATIONS

The presented study will be conducted in accordance with the ICH Harmonised Tripartite Guidelines for Good Clinical Practice and with the ethical principles laid down in the Declaration of Helsinki. The research and innovation activities will comply with ethical principles and relevant national, Union and international legislation, including the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights and its Supplementary Protocols.

Only pseudonymised samples and data will be used for the study. As only existing serum and tissue samples with positive Biobank IC are used, there is a clear consent of the patients for the use of the samples for research purposes.

Funded by the European Union

4.18 DATA PROCESSING AND ACCESS

All data as part of this project will be subject to local data protection laws. Internally, a Biobank PID will be assigned to all samples. Both serum and cryo samples will be processed pseudonymously. The clinical data of the study are fully available in the Clinical Division of Oncology and the Institute of Medical Informatics of the Medical University of Graz and are processed for scientific evaluation by these two institutions. The data remains permanently stored in the database of the test institute.

Data Management

The data preparation will be conducted by the Institute for Medical Informatics and the Division of Oncology. The technical coordinator will receive the data in a pseudonymised form. Data storage management and data processing nodes for AI modules will be set up, utilising the ARIS HPC system with Louros DC operated by GRNET S.A. (National Infrastructures for Research and Technology S.A.) as the central node, endorsing EU-based green data centers.

4.19 STUDY DURATION AND FUNDING

The DIOPTRA project duration is 4 years and it is funded by the European Commission (Project: 101096649 — DIOPTRA — HORIZON-MISS-2021-CANCER-02)

4.20 ANNEX

GRANT AGREEMENT: Project 101096649 — DIOPTRA

4.21 LITERATURE

[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.2021.101174.

[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] R. H. Henderson et al., “The economic burden of colorectal cancer across Europe: a populationbased costof-illness study,” Lancet Gastroenterol. Hepatol., vol. 6, no. 9, pp. 709–722, Sep. 2021, doi:

10.1016/S2468-1253(21)00147-3.

[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] A. Gini et al., "Impact of colorectal cancer screening on cancer-specific mortality in Europe: A systematic review," Eur. J. Cancer, vol. 127, pp. 224–235, Mar. 2020, doi: 10.1016/j.ejca.2019.12.014.

[6] 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.

[7] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era," Curr. Oncol., vol. 28, no. 3, pp. 1581–1607, Apr. 2021, doi: 10.3390/curroncol28030149.

[8] S. Li, L.-A. Miller-Wilson, H. Guo, Y. Peng, and D. A. Fisher, "S239 Adherence to Colorectal Cancer Screening and Associated Healthcare Resource Utilization, a Longitudinal Analysis in US Medicare Population with Ten Years Follow-up," Off. J. Am. Coll. Gastroenterol. ACG, vol. 116, p. S106, Oct. 2021, doi: 10.14309/01.ajg.0000773428.24151.82.

Funded by the European Union

5 PROSPECTIVE CLINICAL STUDY PROTOCOL

This section presents the Prospective Study protocol. After submitting it to the bioethics committees of their countries, partners will receive bioethics permissions.

Prospective study protocol	
Protocol number	1
Document version	1.0
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 it may not be further disclosed by them.

_____ _____

(position, name, surname) (signature)

Ethical principles Statement

The information contained in this document and in related documents, which are prepared for this study, is in accordance with the recognised ethical principles (that 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, 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 have been obtained.

_____ _____

(position, name, surname) (signature)

Approval sheet

_____ _____

(position) (name, surname)

_____ _____

(signature) (date)

Funded by the European Union

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			

5.1 DOCUMENT HISTORY

Revision	Date of enactment	Change author	Change description
1.0			Document is created

5.2 ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
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
BLOCKS	Blocks Health and Social Care EOOD
BURGOS	Fundacion Burgos Por La Investigacion De La Salud
CHUL	Centre Hospitalier Universitaire De Liege
CRC	Colorectal Cancer
FU	Follow Up
GOC	Linac-Pet Scan Opco Limited
NKUA	National and Kapodistrian University of Athens
PSD	Prospective Study Design
RM-RRH	Region Midtjylland
SOP	Standard Operating Procedure
UKCM	Univerzitetni Klinicni Center Maribor

Funded by the European Union

5.3 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 the biological sampling and analysis using in vitro diagnostics.
Purpose	The purpose of this study is the clinical refinement and validation 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 the detection of advanced adenomas. 2. Refinement of the DIOPTRA screening system. 3. Evaluation of the effectiveness of behavioural suggestions to reduce CRC risk. 4. Assessment of cost-effectiveness of DIOPTRA system.
PSD Primary endpoint	Acceptable diagnostic specificity and sensitivity for CRC detection and for the detection of healthy and non-advanced adenoma groups, respectively.
PSD Secondary endpoints	Secondary endpoints include: 1. Acceptable diagnostic sensitivity for the detection of advanced adenomas. 2. Improvement of the performance metrics of the DIOPTRA screening system using the prospective data for refinement. 3. Statistically significant differences in risk factors and protein biomarker concentrations for individuals who have implemented the behavioural suggestions. 4. 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 enrolment.
Subject population	Individuals that visit the clinical sites for a colonoscopy.
Number of subjects	At least 1612 participants are estimated to be recruited in 8 clinical sites (BLOCKS, CHUL, RM-RRH, 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) BLOCKS

Funded by the European Union

	2) CHUL 3) RM-RRH 4) UKCM 5) BURGOS 6) NKUA 7) GOC 8) AGSAVVAS
Prospective Study Procedures	The main study procedures (observational) are the following: 1. Enrolment 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 enrolment, participants will be given the option to potentially participate in the DIOPTRA follow up study. The procedures of the follow up study are: 1. Enrolment in the follow up study after subject eligibility is confirmed. 2. Download of the DIOPTRA mobile app for implementation of 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 within the research and innovation program of the Horizon Europe under Grant Agreement N° 101096649.
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 should be provided for the travel to and back from the hospital.

5.4 BACKGROUND AND 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 from 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]. Particularly, the 5-year survival rate can reach 90% for stage I diagnosis, being less than 15% for advanced stages [6]. Therefore, routine screening is key for reducing mortality and declining incidence rates since CRC is now considered as 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 appears as a promising new tool for non-invasive, quick and safe assessment [5]. Among all liquid biopsy products, blood-derived proteins seem to constitute the most cost-effective solution judging by resources, sensitivity, and research maturity. On this premise, a vast protein pool has been tested, albeit evidence lacks comparative validation, 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 and confusing architecture of most neural network approaches, 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

comprises the main factor assessed by current guidelines, formulating 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, which are 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 wider 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 in 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 from vigorous validation procedures on the retrospective and prospective data (level of evidence B). The study's retrospective part hypothesised that a set of predictive variables is 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).

5.5 STUDY DESIGN

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

STUDY HYPOTHESIS

The main study hypothesis is that the DIOPTRA screening system has adequate clinical performance for the early diagnosis of CRC and advanced adenomas. An additional hypothesis is that the DIOPTRA system can accurately characterise the risk of an individual 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, along with demographic, behavioural, and clinical data from participants belonging to the DIOPTRA study groups to test and refine the DIOPTRA AI models.

TYPE OF INTERVENTION

Funded by the European Union

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: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 2: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 3: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 4: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 5: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 6: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 7: ……..
Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

Research center No 8: ……..

Address: ……..
Tel: ……..; Fax: …… ; e-mail: ……..
Title of the department (s): ……..
Tel: ……..; Fax: …… ; e-mail: …….. .

5.6 OBJECTIVES

The purpose of this study is the clinical refinement and validation 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 the detection of advanced adenomas.
2. Refinement of the DIOPTRA screening system.
3. Evaluation of the effectiveness of behavioural suggestions to reduce CRC risk.
4. Assessment of cost effectiveness of DIOPTRA system.

5.7 ENDPOINTS

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

Secondary endpoints include:

1. Acceptable diagnostic sensitivity for the detection of advanced adenomas.
2. Improvement of the performance metrics of the DIOPTRA screening system using the prospective data for refinement.
3. Statistically significant differences in risk factors and protein biomarker concentrations for individuals who have implemented the behavioural suggestions.
4. Significant reduction of the estimated DIOPTRA screening system costs compared to screening colonoscopy.

5.8 STUDY POPULATION

Prospective study will cover at least 1600 participants to be recruited across all the study's 8 clinical sites (BLOCKS, CHUL, RM-RRH, UKCM, BURGOS, NKUA, GOC, AGSAVVAS). Study population will cover participants that visit the clinical sites for a colonoscopy. Participants will be split into the following groups following the histopathological analysis of index lesions identified during colonoscopy:

- Healthy: no neoplastic findings after a colonoscopy;

- Non-advanced adenomas;
- Advanced adenomas. Under ESGE 2020 guidelines, the following adenoma should be classified as advanced adenomas: at least 1 adenoma ≥ 10 mm or with high-grade dysplasia or with high % of villous growth pattern, or any serrated polyp ≥ 10 mm or with dysplasia;
- Colorectal cancer CRC stage I, II, and III.

Gender distribution in the incidence of CRC will be taken into account in this study, including as similar rates of male and female participants as possible. However, it must be taken into account that males are 25% more prone to develop CRC in comparison to 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).
- 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

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 allowing to manage mobile terminals (smartphones, smartphone apps, tablets).
- Good coverage of internet connection at home.
- Availability of a smartphone/ tablet (in order to be able to use the app).
- Belonging to the healthy or non-advanced adenoma groups

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
- Recent major abdominal surgery (colectomy) or radiation prior to the recruitment
- Inflammatory bowel diseases

- Polyposis syndrome
- Pregnancy or suspicion of pregnancy
- Colorectal cancer history
- 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)

5.9 SAMPLE SIZE

To evaluate the endpoints of the study (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 is satisfactory to reject the NULL hypothesis.

Funded by the European Union

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 is satisfactory to reject the NULL hypothesis.
The confidence intervals for the diagnostic performance metrics of the study, using the calculated sample sizes, are shown in the table below.

Table 1. 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 = 30%;

Funded by the European Union

The required sample size to reject the NULL hypothesis is N = 68 (N = 52 before taking the dropout rate into account) for each DIOPTRA study group in each of the Case and Control groups. In total, N = 272 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 the implementation of DIOPTRA behavioural modification suggestions.

5.10 STUDY PROCEDURES

5.10.1 Overview

For the main observational study, the study procedures are the following:

1. Enrolment 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 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 for the implementation of 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.

Funded by the European Union

5.10.2 Enrolment

Individuals that 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.

5.10.3 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 Protavio (Annex No. 3).

5.10.4 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 be performed according to 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. During the colonoscopy, index lesions will be biopsied and sent for diagnostic analysis according to each site's clinical standards. The study personnel will be responsible to gather the diagnostic data from the biopsies, following the analysis, and match it to the participant's records and forms. The diagnostic results will be used to assign the participants into the DIOPTRA study groups.

5.10.5 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, along with symptoms will be collected during the study via the "Medical Information / History Case Form" (Annex No 5). All collected DIOPTRA data during the study will be uploaded by the study personnel to the DIOPTRA prospective platform.

5.10.6 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. An overview of the study schedule is provided in Table 2, followed by all clinical sites.

Funded by the European Union

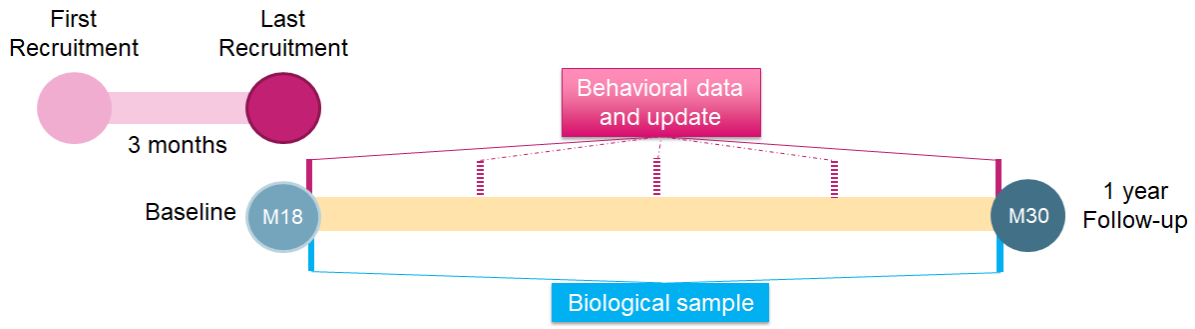
Table 2. Overview of the study schedule

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 end of study.
Follow-up study	Initiation of Follow-up study approximately after 18 months of initial ethical approval.
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	1 year after follow-up enrollment

5.10.7 Follow-up Study Timing and Procedures

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

Scheme 1. Follow-up study timeline



5.10.7.1 Participant Recruitment (M18)

Table 3. Participant recruitment from each clinical site

Participants/clinical site	Healthy	Non-advanced Adenomas
Case (Suggestions)	10	10
Control (No suggestions)	10	10

Prospective cohort enrolment will include participants that correspond to inclusion and exclusion criteria for the follow-up study. After 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 of the DIOPTRA study subgroups: healthy and non-advanced adenoma. In total, 320 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.

5.10.7.2 Behavioural Data Collection (M18, M30 and in-between)

All participants will answer the baseline behavioural data questionnaire in the DIOPTRA mobile app at the moment of recruitment or at the first moment that the participant will be available to answer the questionnaire after the recruitment. A modified follow-up questionnaire will be answered at 1-year follow-up at M12. 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.

5.10.7.3 Biological Sample Collection (M18 and M30)

Blood samples will be collected from each participant at baseline M18 and at the end of the study M30. Samples will be processed and sent to Protavio for biomarker analysis. Details of sample processing could be found in *Annex 2,* with an additional Sample Collection Form in Annex 4.

5.10.7.4 Possible Participant Self-reported Symptoms (M18-M30)

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.

5.10.8 Data Flow and Data Processing (prospective, follow-up)

5.10.8.1 DIOPTRA Software Components

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

1. **Anonymisation Tool (EHR data)**

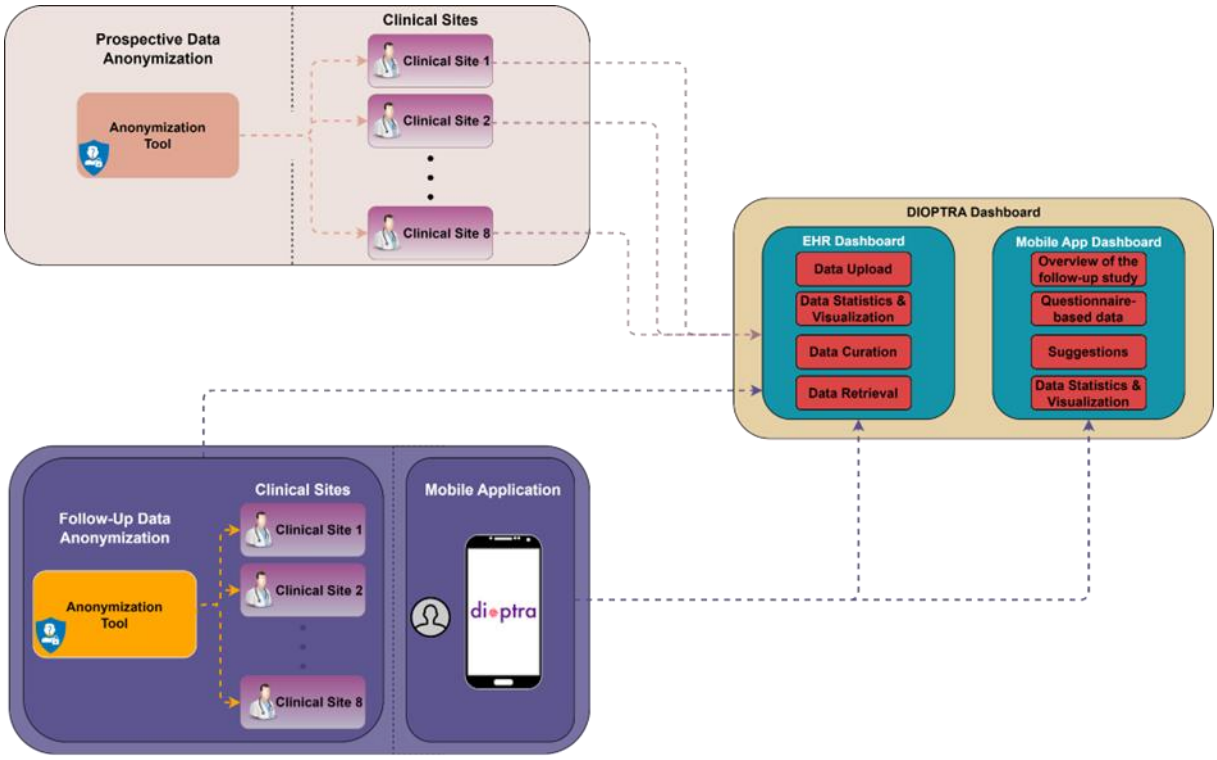
The Anonymisation Tool provides an extra layer of privacy protection to already pseudonymised structured medical data. It will be used for EHR data prior to uploading to the Data Curation & Storage System

2. **Clinical Site Interface**

The clinical sites' DIOPTRA application provides functionality as follows:

EHR Dashboard (data from clinical variables lists) for uploading/downloading tabular retrospective and prospective data in various predefined formats and for providing information about the volume and quality of data uploaded from the clinical site. *Questionnaire Dashboard* providing an overview of follow-up study participants' data collected via the mobile app during the prospective studies and the option to download for further analysis depicted in Figure 1.

Figure 1. Data Curation & Storage Platform (including programming environment)



The anonymised retrospective and prospective data will be stored in the DIOPTRA centralised platform by utilising the ELK STACK. The ELK stack is the main data infrastructure responsible for data collection, curation, and storage, as well as advanced analysis-visualisation (Kibana) and providing M2M data access endpoints for the various data-consuming applications. It offers advanced data

Funded by the European Union

storage and management services, such as: Heterogeneous Data Integration/Ingestion, Data Filtering/Harmonisation, Annotation, Cataloguing and Management, (Meta)Data Storage and Distributed Data Lake, **S**ecurity, Data Protection, Secure Data Sharing, Customised Dashboards and diverse data visualisations, Federated Data Search, retrieval & Interoperable Access.

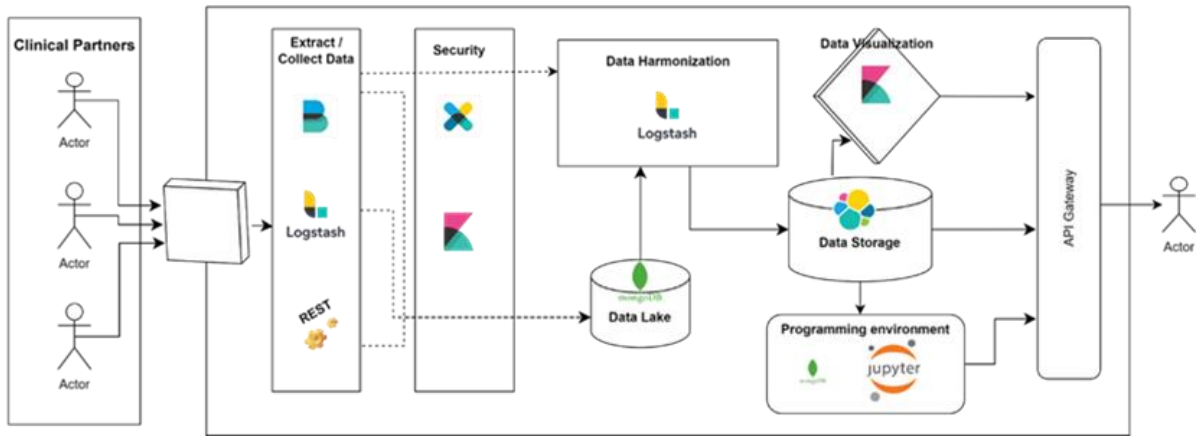
Elasticsearch comes up with Cross-cluster replication (CCR), a way to automatically synchronise indices from the primary cluster to a secondary remote cluster that can serve as backup. If the primary cluster fails, the secondary cluster can take over. Moreover, Elasticsearch provides snapshots as a backup of a running Elasticsearch cluster for data recovery stored in an off-cluster storage location called a snapshot repository.

Data curation techniques will be applied on the retrospective and prospective EHR data. These types of data will be stored in the centralised platform uploaded by each clinical partner. On ensuring overall interoperability and addressing heterogeneity and lack of shared semantics across sources, DIOPTRA will leverage widely adopted ontologies and standards, developing extensions to model relevant knowledge in the domains of the project for which no standards exist.

The Programming Environment provides access to a defined sub-dataset and tools that could be utilised for analysis and/or pattern recognition and model development.

A high-level scheme of the data management architecture is shown below *(Scheme 2).*

Scheme 2. Scheme of the data management architecture



3. **Mobile App**

The mobile application, among other functionality (knowledge base, recommendations, etc.), collects and uploads questionnaire data from follow-up study participants. Each participant's credentials will consist of a unique participant ID and password. The participant must complete a questionnaire based on sociodemographic, 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

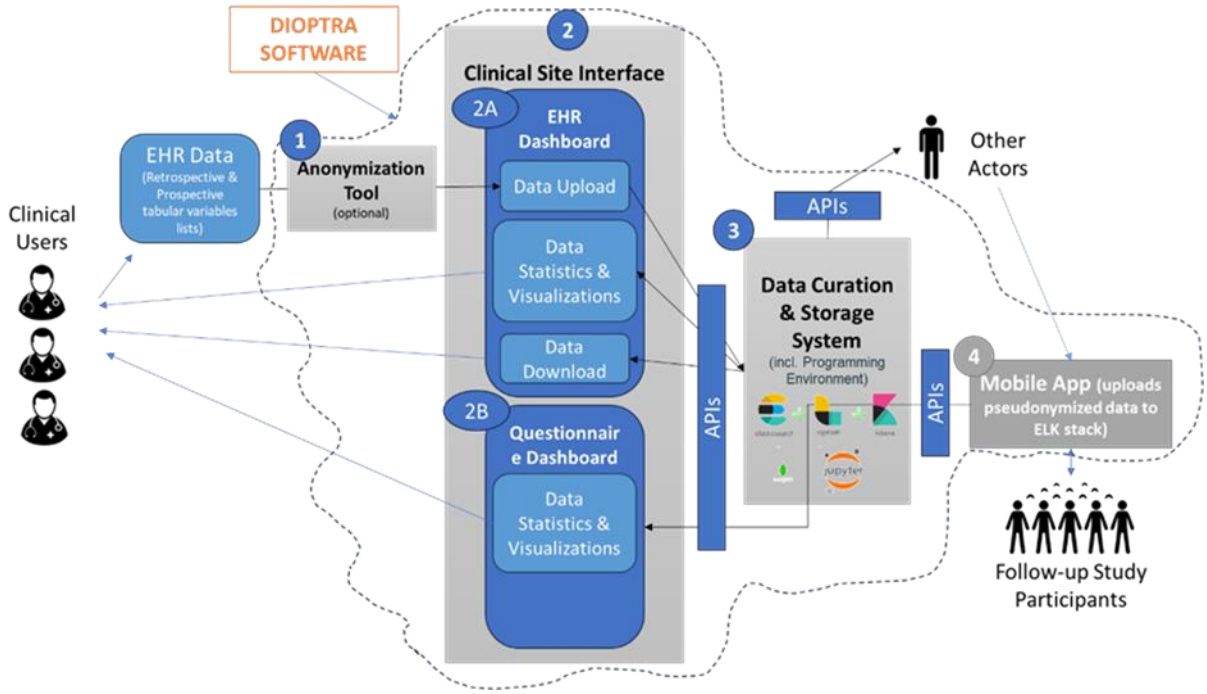
Funded by the European Union

questionnaire data along with the provided suggestions will be stored in the DIOPTRA central storage platform and will be available for review by the clinical users via the Questionnaire Dashboard

5.10.8.2 DIOPTRA Software Data Flow

A conceptual scheme depicting data flow from and to the clinical sites' DIOPTRA application is shown in Scheme 3.

Scheme 3. Data flow from and to the clinical sites' DIOPTRA application



5.10.8.3 Hosting Infrastructure

For data storage, the infrastructure of GRNET will be used by the project. 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

5.10.9 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), 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 techniques can help identify the most informative biomarkers and risk factors for efficient and accurate screening. On the other hand, Deep Learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) will be employed. CNNs can analyse protein biomarker data, while RNNs can process sequential risk factor data to capture temporal patterns and dependencies. Hybrid models that combine deep learning with traditional ML algorithms can offer enhanced performance.
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 4*) for the different endpoints and models will be constructed. Diagnostic metrics will be calculated as Sensitivity = TP/(TP+FN) and Specificity = TN/(TN+FP). To compare the performance metrics to their respective lower bounds the exact binomial test will be utilised. 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 4. 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)

5.11 ETHICS AND DATA MANAGEMENT

Participation in DIOPTRA Prospective Study will be done on the basis of an informed consent that will allow for the voluntary participation in the study and for the processing of personal information of the subjects.

Revision in Patient Information and Informed Consent Form. The organisation responsible for the study will inform the investigator whenever information becomes available that may be relevant to the subject's confirmed participation in the study. The investigator or his/her authorised designee should inform the subject in a timely manner.
The organisation responsible for the study will revise the written Informed Consent Form whenever new information becomes available that may be relevant to the subject's confirmed participation in the study. The revised information 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.

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.

5.12 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

Funded by the European Union

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 accuracy, completeness, and timeliness of the data. Data which 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, as well as any imaging that is sent to the oganisation 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 closure of the study. The frequency and timing of monitoring visits shall be determined by the organisation responsible for the study for each site based on the scope of collected data, study compliance, and findings from previous visits.

The monitoring strategy covers below mentioned actions (*Table 5*).

Table 5. 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 conducts as planned (in full scope and related time frames) and that all changes are well managed.
Monitoring visits: interim visits. Not less than once per 3 months.	Organisation responsible for the study and study team members	Onsite/remote monitoring visits could be conducted.	Interim Monitoring Visits may be conducted throughout the study to verify that: ▪ The clinical site is conducting the study in accordance with 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 the suspension or termination of the protocol procedures.

In this case, the clinical site must inform the Organisation responsible for the study of the reasons for the termination of the study, and the data collected prior to the termination of the study must be passed on to the organisation responsible for the study.
Any changes will be agreed in advance 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.
Organisation responsible for the study will notify the Bioethical Committee about the study closure by providing a Prospective Study report based on the Bioethical Committee/ other regulatory authority form.

5.13 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.

Funded by the European Union

13. 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, a colorectal cancer develops in the colon, also known as the large intestine, or in its last part, the rectum. The starting point of a 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 tumor. The evolution of precancerous polyps into tumors 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 fecal occult blood test, additional test 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 positive fecal 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 a 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 largely unknown, it has been suggested that environmental and behavioral 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.

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

Funded by the European Union

*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 is a 4-year project funded by the European Commission under the Horizon Europe programme with project number 101096649 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 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.), behavioral information (cigarettes, alcohol) and nutritional habits, physical activity, etc.

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 only on the day of your colonoscopy visit:

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 socio-demographic, behavioural information, habits and physical activities etc, which will be asked by a health professional [adjust to local method].

C. Accept that part of your data will also be used anonymously 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 [clinical site].

In order to participate in this study, the medical staff will make sure that you:

- are between 18 and 80 years of age at the time of recruitment
- have a prescription for a total colonoscopy
- do not have a severe systemic abnormality
- are able to understand the study instructions and sign an informed consent form
- have not undergone major abdominal surgery (e.g. colectomy) or radiation treatment prior to colonoscopy
- are not pregnant
- have not been diagnosed with chronic inflammatory bowel disease (e.g. Crohn's disease or ulcerative colitis), polyposis syndrome or colorectal cancer

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, 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 can 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 or if you 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 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*] and other related laws and regulations.

Funded by the European Union

Who is your Data Controller?

[*Organisation responsible for the study*] is the Data Controller of 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 only by the study team of [*organisation responsible for the study*].* After the anonymisation of your personal data, the study team may share these anonymised data with other organisations participating in the DIOPTRA project for the purpose of the study. 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?

No, your data will not be transferred to anyone outside the European Union/European Economic Area.

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.

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

We aim to anonymise your personal data upon collection. However, your personal data that has not been anonymised will be stored in the local servers of the Data Controller, and 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 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, 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, and the study team has already anonymised your data for the study, they are still allowed to use this anonymised information as it does not contain any personal data. 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 calmly decide if I wanted to take part.

- 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:

Funded by the European Union

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)										
								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.

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. 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 tumor. The evolution of precancerous polyps into tumors 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**

Funded by the European Union

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 already symptomatic stage of the disease. Therefore, although more invasive, colonoscopy remains the most reliable method of screening for colorectal cancer, as it enables polyps and other lesions to be visualised and removed using an endoscope 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 a tumor. 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 behavioral 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 is a 4-year project funded by the European Commission under the Horizon Europe programme with project number 101096649, 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 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.

Funded by the European Union

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.), behavioral 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 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. Accept that part of your data will also be used anonymously 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 [clinical site].

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

- are between 18 and 80 years of age at the time of recruitment
- have a prescription for a total colonoscopy
- do not have a severe systemic abnormality
- are able to understand the study instructions and sign an informed consent form
- have not undergone major abdominal surgery (e.g. colectomy) or radiation treatment prior to colonoscopy

Funded by the European Union

- are not pregnant

- have not been diagnosed with chronic inflammatory bowel disease (e.g. Crohn's disease or ulcerative colitis), polyposis syndrome or colorectal cancer

Phase II:

You will participate in this phase:

A. Only if your colonoscopy shows no signs of colon cancer, which we hope you will.

B. If you have sufficient technological knowledge to manage smartphones and smartphone applications and agree to use the DIOPTRA application

C. Have good Internet connection coverage in your place of residence.

This second phase of the study will involve 320 participants, including 40 from Hospital.

After your hospital visit in Phase I, you will be contacted on 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 reduce 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 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 be also 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.

Funded by the European Union

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, 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.

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.

Funded by the European Union

Your participation in the study will be automatically terminated if you no longer belong to the subject inclusion criteria or if you 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 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 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 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, your 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, your answers to the questionnaires provided at the clinical site as well as in the mobile app related to demographic, dietary, financial, lifestyle and behavioural data 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 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 only by the study team of [organisation responsible for the study]. After the anonymisation of your personal data, the study team may share these anonymised data with other organisations participating in the DIOPTRA project for the purpose of the project. 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?

No, your data will not be transferred to anyone outside the European Union/European Economic Area.

How do we protect your privacy?

To protect your personal data, your identity information will be replaced by a code. 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.

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

We aim to anonymise your personal data upon collection. However, your personal data that has not been anonymised will be stored in the local servers of the Data Controller, and 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 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 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 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, and the study team has already anonymised your data for the study, they are still allowed to use this anonymised information as it does not contain any personal data. 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 calmly decide if I wanted to take part.

Funded by the European Union

- 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 be contacted by the study team to participate in the second phase of the study after my first colonoscopy visit.	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.

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 that you took the most part 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?	☐ High school graduate ☐ Middle school graduate ☐ College or University degree ☐ Post-graduate degree
What is your monthly net income?	☐400€ to 1000€ ☐Less than 1000€ ☐1000€ to 2.000€ ☐2000€ to 3000€ ☐ More than 3000€
What is your occupation?	☐ Professional/Technical/Scientific ☐ Healthcare/Medical ☐ Managerial/Supervisory ☐ Education/Academic ☐ Business/Entrepreneur ☐ Arts/Creative ☐ Public Service/Government ☐ Retail/Hospitality ☐ Sales/Customer services ☐ Administrative/Clerical ☐ Skilled Trades/Manual Labor ☐ Student/not currently employed ☐ Retired ☐ Homemaker ☐ Unemployed ☐ Other

[1]This refers to areas characterised 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.

[1]Professional occupations in natural and applied sciences, health, education, law and social, community and government services

[2]Management, business, finance, and administration occupations

[3]Sales and service occupations, including occupations related to the hospitality and tourism industries

[4]Administrative and office support occupations

[5]Occupations in manufacturing (e.g., metal, glass, chemicals, wood, pulp, textile), agriculture and

natural resources (e.g., farming, fishing, forestry), construction, trades, transport, and equipment operation

Lifestyle (mark with an x)	
Are you smoking?	☐ Yes ☐ No
On average, how many cigarette packs[1] do you smoke per day?	☐ Less than 1 ☐ 1 to 2 ☐ More than 2
If you are currently smoking, how long have you been smoking? (Write a number in years)	
If you have stopped smoking, how long have you been smoking in the past? (Write a number in years)	
Are you regularly exposed to secondhand smoke[2]?	☐ Yes ☐ No
How many days per week do you consume alcohol[3]?	☐ 0 to 1 ☐ 1 to 3 ☐ 4 to 6 ☐ 7
How many standard alcoholic drinks[3] 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 90 min of moderate physical activity per week) ☐ Moderately active (90 to 150 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 30 minutes ☐ 30 to 60 minutes ☐ More than 60 minutes
On average, how many days per week do you engage in physical activity? (Write in days)	☐ Less than 2 days ☐ 2 to 4 days ☐ More than 4 days

Funded by the European Union

How long do your typical physical activity sessions last? (Write in minutes)	
What is your average daily sedentary time[4]?	☐ Less than 5 hours ☐ 5 to 10 hours ☐ More than 10 hours
How many hours per day do you engage in prolonged sitting[4]?	☐ Less than 2 hours ☐ More than 2 hours
[1]Suppose that 1 pack includes 20 cigarettes [2]Daily exposure to the tobacco smoke of others at home, work, or public places [3]Frequency of drinking alcoholic beverages (e.g., 354 ml can/bottle of beer, 118ml glass of wine, 44ml shot of hard liquor [4]. 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, socialising)). It includes any time spent in activities with minimal energy expenditure. [5]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

Funded by the European Union

[1] Examples of fruit: fresh fruit, chopped, cooked or canned fruit, dried fruit, fruit juice. Examples of vegetables: raw leafy vegetables, cooked, canned, frozen, or chopped vegetables, 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 specialty 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., burger, fries, taco), pizza, and instant meals (e.g., instant ramen noodles)

Supplements (mark with an x)	
How often do you consume omega 3 (including multivitamin)?	☐ Never[1] ☐ Rarely[2] ☐ Often[3]
Do you take a daily multivitamin supplement?	☐ Yes ☐ No
If not, which supplements do you consume?	☐ None of them ☐ Vitamin B6 ☐ Vitamin C ☐ Vitamin D ☐ Magnesium ☐ Calcium
How often do you consume these supplements (answer of previous question)?	☐ Never ☐ Rarely ☐ Often
How often do you consume probiotics[4]?	☐ Never ☐ Rarely ☐ Often
How often do you consume fiber 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]Never: Indicates that the supplement is not consumed at all.
[2]Rarely: Indicates that the supplement is consumed infrequently or occasionally, but not on a regular basis.
[3]Often: Indicates that the supplement is consumed frequently or regularly as part of the dietary routine.
[4] 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. SAMLE COLLECTION & MANAGEMENT

1. SCOPE OF THE PROCEDURE

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

a. the collection, labeling, 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 (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 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
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

Funded by the European Union

3.2 MATERIALS

#	Description	Specifications	Recommended Cat No
1	9-10mL Serum collection tubes	Plastic, 16x100mm, with clot activator (silica), **red cap color**, transparent	BD Vacutainer, #367896 Greiner Vacuette, #455092
2	9-10mL K2EDTA Plasma collection tubes	Plastic, 16x100mm, with K2EDTA additive, **purple/lavender cap color**, 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 anonymisation of personal data during the submission of samples to the Test Facility performing the biological analysis.

The following identification system **should be followed** for a codification of participants:
Each ID will include the clinical site code followed by a 4-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
BLOCKS	CP09
CHUL	CP12
RM-RRH	CP15
UKCM	CP16

Funded by the European Union

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 aliquot*.
Example: CP24-0034-S-1

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

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

4.3 TUBE LABELING

Each storage tube should be clearly labeled 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 of the 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 blood draw in the Collection Form.
2. Gently invert tube 5-6 times to mix blood with clot activator.

3. Place upright on a test rack and allow to sit for 30-60 min at 18-25°C until clotting has occurred.
4. Centrifuge at 1,500-2,000 x g for 10 minutes at 18-25°C. 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 -80°C 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 18-25°C.
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 cryoboxes in an ultra-low freezer at -80°C 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

Funded by the European Union

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-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 in to the Study Director.

Funded by the European Union

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 labeling on the test item matches information recorded by the organisation responsible for 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 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°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
- No of tubes received
- Volume per tube
- Total volume received

Funded by the European Union

- Visual Inspection results (normal, hemolytic, icteric or lipemic)
- Storage Temperature
- Storage Location
- Box ID
- Comments/Deviations from receipt process
- Comments/Deviations during sample collection (upon inspection of Sample Collection Form by the clinical partner)

10. CLINICAL CHARACTERISATION

The clinical information of each participant will be collected in the Medical Information/History Case Form and will be used to categorise 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 from each Clinical Partner two aliquots of serum and two aliquots of plasma samples per participant. Each Clinical Partner will retain two aliquots of serum and two aliquots of plasma samples per participant as back-up 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 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 REPSONSIBLE FULL NAME	
COLLECTION REPSONSIBLE 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)	
Withdrawal time (min)	
Time required for interventions (min)	
Total Procedure Time (min)	
Gloucester Comfort Score (1-5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Funded by the European Union

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 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	☐ Hemolysed (red colour)
	☐ Icteric (bright yellow)	☐ Lipemic (turbid)
Comments – Deviations from SOP Different Tube type, 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	

Funded by the European Union

Number of storage tubes prepared (0.5 mL per tube)		
Left-over volume (mL)		
Storage Temperature (^{O}C)		
Visual Inspection	☐ Normal	☐ Hemolysed (red colour)
	☐ Icteric (bright yellow)	☐ Lipemic (turbid)
Comments – Deviations from SOP Different Tube type, longer waiting times, low sample volumes, other centrifugation conditions etc		

Funded by the European Union

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 ☐ 2nd degree ☐ No
What was the age of the youngest relative at diagnosis?	☐ Less than 50 years ☐ More than 50 years ☐ Unknown
What is the sex of the relative?	☐ Male ☐ Female

Personal History (mark with an x)	
Have you had a previous colonoscopy?	☐ Yes ☐ No
If yes, how many years ago?	
What were the findings of your last colonoscopy?	☐ Healthy ☐ Non advanced adenoma ☐ Advanced adenoma
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 ☐ No
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

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 ☐ NSAID ☐ 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 ☐ 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 anemia?	☐ Yes ☐ No

Female 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
Have you ever used oral contraceptive?	☐ Never ☐ Rarely ☐ Sometimes ☐ Fairly often ☐ Very often
What is your current menopausal status?	☐ Pre-menopausal ☐ Post-menopausal

Funded by the European Union

6 CONCLUSIONS

This deliverable outlines the protocols governing the DIOPTRA clinical studies conducted at participating clinical sites, including all related documents such as questionnaires, consent forms, etc.

Regular monitoring and evaluation of the protocols will be ensured throughout the implementation of the studies across all sites, assessing the prepared protocols' effectiveness with regard to the project's goals. In this regard, the content will be adjusted as needed, while the updated (and final) versions of all clinical protocols of DIOPTRA will be included in a future project deliverable (D6.3: Updated DIOPTRA Clinical Protocols).

Funded by the European Union

REFERENCES

[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.2021.101174

[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] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," CA. Cancer J. Clin., vol. 70, no. 1, pp. 7– 30, 2020, doi: 10.3322/caac.21590.

[5] A. Gini et al., "Impact of colorectal cancer screening on cancer-specific mortality in Europe: A systematic review," Eur. J. Cancer, vol. 127, pp. 224–235, Mar. 2020, doi: 10.1016/j.ejca.2019.12.014.

[6] 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.

[7] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era," Curr. Oncol., vol. 28, no. 3, pp. 1581– 1607, Apr. 2021, doi: 10.3390/curroncol28030149.

[8] 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.

[9] K. Simon, "Colorectal cancer development and advances in screening," Clin. Interv. Aging, vol. 11, pp. 967– 976, Jul. 2016, doi: 10.2147/CIA.S109285.

[10] S. Li, L.-A. Miller-Wilson, H. Guo, Y. Peng, and D. A. Fisher, "S239 Adherence to Colorectal Cancer Screening and Associated Healthcare Resource Utilization, a Longitudinal Analysis in US Medicare Population with Ten Years Follow-up," Off. J. Am. Coll. Gastroenterol. ACG, vol. 116, p. S106, Oct. 2021, doi: 10.14309/01.ajg.0000773428.24151.82.

[11] "Colorectal Screening In Europe," Digestive Cancers Europe. https://digestivecancers.eu/publication/colorectal-screening-in-europe/ (accessed Apr. 12, 2022).

[12] 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: Feb. 17, 2023. [Online]. Available: https://www.frontiersin.org/article/10.3389/fcell.2021.660924

Funded by the European Union

[13] 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.

[14] J. L. Petrick, L. E. Barber, S. Warren Andersen, A. A. Florio, J. R. Palmer, and L. Rosenberg, "Racial Disparities and Sex Differences in Early- and Late-Onset Colorectal Cancer Incidence, 2001–2018," Front. Oncol., vol. 11, 2021, Accessed: Feb. 17, 2023. [Online]. Available: https://www.frontiersin.org/article/10.3389/fonc.2021.734998

[15] A. S. Butterworth, J. P. T. Higgins, and P. Pharoah, "Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis," Eur. J. Cancer Oxf. Engl. 1990, vol. 42, no. 2, pp. 216– 227, Jan. 2006, doi: 10.1016/j.ejca.2005.09.023.

[16] T. Jess, C. Rungoe, and L. Peyrin-Biroulet, "Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies," Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc., vol. 10, no. 6, pp. 639–645, Jun. 2012, doi: 10.1016/j.cgh.2012.01.010.

[17] Y. Dong et al., "Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies," Biosci. Rep., vol. 37, no. 6, p. BSR20170945, Dec. 2017, doi: 10.1042/BSR20170945.

[18] E. L. Amitay et al., "Smoking, alcohol consumption and colorectal cancer risk by molecular pathological subtypes and pathways," Br. J. Cancer, vol. 122, no. 11, Art. no. 11, May 2020, doi: 10.1038/s41416-020- 0803-0.

[19] H. Kim et al., "Total Vitamin D Intake and Risks of Early-Onset Colorectal Cancer and Precursors," Gastroenterology, vol. 161, no. 4, pp. 1208-1217.e9, Oct. 2021, doi: 10.1053/j.gastro.2021.07.002.

[20] M. C. Hornbrook et al., "Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data," Dig. Dis. Sci., vol. 62, no. 10, pp. 2719–2727, Oct. 2017, doi: 10.1007/s10620-017-4722-8.

[21] J. D. Warren et al., "Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer," BMC Med., vol. 9, no. 1, pp. 1–9, Dec. 2011, doi: 10.1186/1741-7015-9-133/COMMENTS.

[22] S. Harlid, M. J. Gunter, and B. Van Guelpen, "Risk-Predictive and Diagnostic Biomarkers for Colorectal Cancer; a Systematic Review of Studies Using Pre-Diagnostic Blood Samples Collected in Prospective Cohorts and Screening Settings," Cancers 2021 Vol 13 Page 4406, vol. 13, no. 17, p. 4406, Aug. 2021, doi: 10.3390/CANCERS13174406.

[23] R. Meng, Y. Wang, L. He, Y. He, and Z. Du, "Potential diagnostic value of serum p53 antibody for detecting colorectal cancer: A meta-analysis," Oncol. Lett., vol. 15, no. 4, pp. 5489–5496, Apr. 2018, doi: 10.3892/OL.2018.8070/HTML.

[24] N. Hauptman and D. Glavač, "Colorectal Cancer Blood-Based Biomarkers," Gastroenterol. Res. Pract., vol. 2017, p. 2195361, 2017, doi: 10.1155/2017/2195361.

[25] C. Dai et al., "Serum macrophage inhibitory cytokine-1 serves as a novel diagnostic biomarker of early-stage colorectal cancer," Biomark. Biochem. Indic. Expo. Response Susceptibility Chem., vol. 26, no. 7, pp. 598– 605, 2021, doi: 10.1080/1354750X.2021.1950209.

[26] David A. Brown; et al., "MIC-1 Serum Level and Genotype | Clinical Cancer Research | American Association for Cancer Research," Clin Cancer Res, vol. 9, no. 7, pp. 2642–2650, 2003.

[27] X. Wang et al., "Circulating MIC-1/GDF15 is a complementary screening biomarker with CEA and correlates with liver metastasis and poor survival in colorectal cancer," Oncotarget, vol. 8, no. 15, pp. 24892– 24901, Feb. 2017, doi: 10.18632/ONCOTARGET.15279.

[28] D. Ding, S. Han, H. Zhang, Y. He, and Y. Li, "Predictive biomarkers of colorectal cancer," Comput. Biol. Chem., vol. 83, p. 107106, Dec. 2019, doi: 10.1016/J.COMPBIOLCHEM.2019.107106.

[29] A. Herbst et al., "Methylation of NEUROG1 in serum is a sensitive marker for the detection of early colorectal cancer," Am. J. Gastroenterol., vol. 106, no. 6, pp. 1110–1118, Jun. 2011, doi: 10.1038/ajg.2011.6.

[30] 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.

[31] 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.

[32] N. Laprovitera, M. Riefolo, E. Ambrosini, C. Klec, M. Pichler, and M. Ferracin, "Cancer of Unknown Primary: Challenges and Progress in Clinical Management," Cancers, vol. 13, no. 3, p. 451, Jan. 2021, doi: 10.3390/cancers13030451.

[33] 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.

[34] 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.

[35] 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.

[36] P. Wang et al., "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study," Gut, vol. 68, no. 10, pp. 1813–1819, Oct. 2019, doi: 10.1136/gutjnl-2018-317500.

Funded by the European Union

[37] F. Soares, K. Becker, and M. J. Anzanello, “A hierarchical classifier based on human blood plasma fluorescence for non-invasive colorectal cancer screening,” Artif. Intell. Med., vol. 82, pp. 1–10, Oct. 2017, doi: 10.1016/j.artmed.2017.09.004.

[38] A. Nogueira-Rodríguez et al., “Deep Neural Networks approaches for detecting and classifying colorectal polyps,” Neurocomputing, vol. 423, pp. 721–734, Jan. 2021, doi: 10.1016/j.neucom.2020.02.123.

[39] Y. Kinar et al., “Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study,” J. Am. Med. Inform. Assoc. JAMIA, vol. 23, no. 5, pp. 879–890, Sep. 2016, doi: 10.1093/jamia/ocv195.

[40] M. Y. Lu et al., “AI-based pathology predicts origins for cancers of unknown primary,” Nature, vol. 594, no. 7861, pp. 106–110, Jun. 2021, doi: 10.1038/s41586-021-03512-4.

[41] M. F. Mohamad Marzuki, N. A. Yaacob, N. M. bin Yaacob, M. R. Abu Hassan, and S. B. Ahmad, “Usable Mobile App for Community Education on Colorectal Cancer: Development Process and Usability Study,” JMIR Hum. Factors, vol. 6, no. 2, p. e12103, Apr. 2019, doi: 10.2196/12103.

[42] 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.

[43] 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.

Funded by the European Union