

# HEAL ITALIA Health Extended Alliance for Innovative Therapies, Advanced Lab-research, and Integrated Approaches of Precision Medicine

**Tematica** Medicina di precisione

**Referente Univpm** Prof. Gianluca Moroncini

**Soggetto attuatore** Fondazione "HEAL ITALIA"

**Costo progetto** € 118.765.450,00

**Finanziamento PNRR progetto** € 114.700.000,00

**Ruolo Univpm** Spoke Leader 7

Affiliato allo Spoke 1 (TOR VERGATA)

Affiliato allo Spoke 3 (UNIPA)

Affiliato allo Spoke 4 (LA SAPIENZA)

**Budget Univpm** € 6.225.000,00

**Sito web:** <https://www.healitalia.eu/>

**Descrizione del progetto di ricerca svolto da UNIVPM:**

**SPOKE - 7**

**STRATEGIE DI PREVENZIONE**

**Approcci di medicina integrata e di genere per strategie di prevenzione basate su dati ambientali, di stile di vita e biometrici clinici**

**Obiettivi generali:**

La prevenzione è un pilastro delle politiche della salute pubblica con lo scopo di ridurre il peso delle malattie. Il livello di implementazione delle strategie preventive è influenzato dalle, ma è anche un determinante delle, condizioni socioeconomiche di un paese. Le strategie di prevenzioni convenzionali non considerano la suscettibilità individuale a sviluppare malattie ma trovano applicazione su larghi gruppi o su un'intera popolazione, e interventi che potrebbero ottenere risultati rilevanti in casi selezionati potrebbero non mostrare efficacia quando adottati su larga scala. Le caratteristiche genetiche individuali o le condizioni biologiche in concomitanza con gli agenti ambientali e le abitudini di vita determinano il rischio di una malattia, e la traduzione delle conoscenze in interventi mirati massimizzerebbe il loro rapporto costo-efficacia. Ci proponiamo di identificare, in un approccio di medicina di genere, i determinanti del rischio individuale di sviluppare malattie ad alto impatto sullo stato di salute di una popolazione, come le malattie cardiovascolari, tumorali o endocrino-metaboliche. L'identificazione dei determinanti del rischio individuale può essere la base di nuove strategie innovative di prevenzione.

**WP1: Traduzione degli strumenti della medicina di precisione in nuove strategie per la prevenzione del cancro**

**Task 1.1: Predizione dello sviluppo del cancro nelle condizioni infiammatorie croniche del sistema digerente: profilazione molecolari, predizione del rischio individuale e potenziali strategie di prevenzione.**

**1.1.1:** Modelli animali specifici di malattie epatiche croniche e strutture in vitro verranno utilizzati per identificare i marcatori della progressione del danno epatico e dello sviluppo del cancro. I dati verranno confrontati con le caratteristiche prognostiche cliniche della progressione della malattia provenienti da database locali, nazionali e internazionali. Specifici fattori ambientali e abitudini di vita verranno valutati come parametri eziologici non trasmissibili in grado di modificare morbilità e mortalità. Verrà sviluppata una piattaforma per la caratterizzazione multigenica, insieme a strumenti a supporto del processo decisionale clinico. Deliverable: Profilo molecolare dei pazienti a rischio di sviluppare tumori epatici primari.

**1.1.2:** Studio del ruolo delle tossine batteriche nella tumorigenesi e nella carcinogenesi del colon-retto, per definire profili genetici associati ad un aumento del rischio e meccanismi attuabili, con l'obiettivo finale di esplorare strategie preventive basate sulla dieta di interventi di modificazione del microbioma. Deliverable: Identificazione dei meccanismi immunitari, metabolici, epigenetici modulati dalle tossine batteriche.

### **Task 1.2: Ponderazione dell'esposoma: un modello per la prevenzione delle malattie**

È stata allestita una BioBanca di campioni biologici di soggetti ex-esposti ad amianto (ex-EXP), di pazienti affetti da mesotelioma pleurico maligno (MPM) e di soggetti sani. Biomarcatori circolanti di diversa origine legati all'esposizione all'amianto, come miRNA correlati all'amianto, biomarcatori autofagici della trasformazione mesoteliale ATG5 e HMGB1, e proteine solubili correlate alla mesotelina (SMRPs) saranno studiati per la predizione e il follow-up del MPM. Deliverables: i) un algoritmo per la prevenzione e la diagnosi precoce delle malattie tumorali legate all'amianto; ii) firme molecolari delle malattie tumorali legate all'amianto e mortalità di follow-up del MPM.

### **Task 1.3: Accelerare lo sviluppo di terapie innovative per i tumori solidi mediante evidenze precliniche avanzate e sistemi 3D**

Dispositivi microfluidici (immune-on-a-chip) generati con le cellule primarie del paziente per analizzare le interazioni di specifiche popolazioni di cellule immunitarie con i tessuti patologici dell'ospite, e per testare l'attività delle immunoterapie o l'effetto delle terapie convenzionali sul sistema immunitario, in modo paziente-specifico, facilitando i percorsi della medicina personalizzata. Deliverable: Identificazione dei segnali immunitari chiave che guidano l'efficacia delle terapie antitumorali.

### **Task 1.4: Medicina predittiva oncologica della testa e del collo mediante analisi omica di intelligenza artificiale**

Identificazione del panorama genomico e molecolare che è alla base della diversità fenotipica nella biopsia liquida dell'HNSCC utilizzando l'analisi trascrittomico a singola cellula e il sequenziamento di prossima generazione, correlando questi risultati a fenotipi specifici mediante approcci di intelligenza artificiale. Deliverables: biomarcatori genomici, trascrittomici e microbiotici per la diagnostica predittiva, la prognosi e lo sviluppo di farmaci nei pazienti con HNSCC.

### **Task 1.5: Predizione del rischio di cancro nella popolazione italiana**

Implementazione di un metodo innovativo e completo per la predizione del rischio di cancro nella popolazione italiana, attraverso la definizione di pannelli completi di geni tumorali (mediante NGS) e la selezione di SNP comuni da genotipizzare per la definizione di punteggi di rischio poligenico (PRSs) per i più frequenti tipi di cancro. In collaborazione con la società "Allelica", saranno sviluppati algoritmi per la valutazione di PRS specifici per il cancro e strumenti per l'analisi integrativa della genetica ad alta/moderata/bassa penetranza e

dell'esposizione ambientale. Deliverables: Fornire modelli efficaci per la stratificazione della popolazione italiana per la stima del rischio di cancro.

## **WP2: Nuove strategie personalizzate per la prevenzione delle malattie cardiovascolari**

### **Task 2.1: Prevenzione della Cardiomiopatia Aritmogena (ACM)**

La cardiomiopatia aritmogena (ACM) è una malattia rara, genetica, ereditaria del muscolo cardiaco che è una delle principali cause di morte cardiaca improvvisa (MCI), soprattutto nei giovani adulti e negli atleti e che viene spesso mal diagnosticata. Con test di laboratorio, registrazioni ECG, risonanza magnetica cardiaca, un'accurata mappatura del voltaggio elettroanatomico, una biopsia endomiocardica e test genetici, ipotizziamo di ridurre la percentuale di diagnosi errate e di identificare meglio le fenocopie. Deliverable: un percorso diagnostico in grado di riconoscere i pazienti con ACM da fenocopie patologiche e uno score multiparametrico che consenta l'identificazione di pazienti ad alto rischio aritmico, utilizzando dati clinici, di imaging, elettrofisiologici, istologici e genetici.

### **Task 2.2: Valutazione dei fattori di rischio per l'insufficienza cardiaca cronica**

Sviluppare e validare un modello adatto per monitorare la densità del  $\beta$ 1AR come possibile marcatore predittivo dello sviluppo di insufficienza cardiaca (HF) utilizzando monociti del sangue periferico isolati da soggetti sani maschi e femmine. Deliverable: identificazione di possibili marcatori predittivi dello scompenso cardiaco.

### **Task 2.3: Un nuovo punteggio di rischio specifico per le donne predittivo di aterosclerosi per strategie di prevenzione personalizzate**

Il danno cardiovascolare (CV) è associato a una percentuale significativa di decessi nelle donne. Le donne sono esposte a specifici fattori di rischio CV legati al sesso, tra cui la menopausa e la depressione, e anche i tradizionali fattori di rischio che agiscono in modo diverso nelle donne rispetto agli uomini. Studieremo tre diverse coorti dalla pre- alla post- menopausa. Verranno utilizzati metodi di apprendimento automatico per integrare dati molecolari, fenotipici e sullo stile di vita per generare un nuovo punteggio di rischio specifico per le donne predittivo dell'aterosclerosi. Deliverable: punteggio di rischio cardiovascolare sesso-specifico pronto e validato.

### **Task 2.4: Nuovi predittori precoci di infezioni di dispositivi cardiovascolari elettronici impiantabili**

L'uso di dispositivi cardiovascolari elettronici impiantabili (CIEDs) è diffuso e ha dimostrato di migliorare la qualità della vita dei pazienti e la sopravvivenza a lungo termine. Proponiamo una caratterizzazione approfondita dei pazienti con sospetta infezione del CIED al fine di i) identificare i fattori di rischio per l'infezione sistemica rispetto a quella locale; ii) identificare i fattori di rischio per la recidiva dell'infezione del CIED o per esiti avversi; iii) fornire nuovi predittori precoci dell'infezione del CIED. Deliverable: un nuovo strumento di stratificazione per prevedere il rischio di reinfezione del CIED dopo la rimozione dell'hardware e creare un nuovo algoritmo per la gestione dei pazienti con CIED dopo la rimozione dell'hardware.

## **WP3: Integrazione dei vecchi fattori di rischio e nuovi modelli predittivi per la prevenzione delle patologie metaboliche ed endocrine**

### **Task 3.1: Fattori di rischio, stile di vita e nuovi biomarcatori nell' obesità e nelle patologie correlate**

Studiare i meccanismi e gli effetti specifici dei nutrienti e dei composti che alterano il metabolismo (MDCs) sul tessuto adiposo e sugli organoidi derivati dalle cellule staminali pluripotenti indotte (iPSC) umane in una corte di pazienti prendendo in considerazione sesso, età, stato di salute, apporto alimentare ed integrazione con altri organi, inclusi intestino e cervello. Verranno utilizzati modelli murini validati per verificare la risposta dell'asse intestino-cervello e i processi infiammatori coinvolti nella regolazione delle patologie metaboliche e i disturbi dell'umore. Deliverable: Caratterizzazione di specifici gruppi a rischio per l'esordio di patologie metaboliche legate ai nutrienti e/o agli effetti dell'MDC.

### **Task 3.2: Diasmoke – Studio clinico randomizzato controllato per valutare i cambiamenti nel rischio cardiovascolare in pazienti con diabete di tipo 2 che passano a sistemi di somministrazione di nicotina senza combustione**

Ampio studio randomizzato di commutazione per indagare i cambiamenti nei fattori di rischio CV e i parametri metabolici nei fumatori diabetici che passano a sistemi di somministrazione di nicotina senza combustione - CF NDS (ad esempio sigarette elettroniche, prodotti a base di tabacco riscaldato), confrontandoli con quelli che continuano a fumare. Deliverables: identificazione dei cambiamenti dei fattori di rischio CV nei pazienti diabetici

### **Task 3.3: Mirare alle funzioni riproduttive per nuove strategie di prevenzione**

Identificazione di nuove strategie di prevenzione e trattamento precoce per le patologie legate alla riproduzione, al fine di assicurare una salute sessuale e riproduttiva, uno dei goals di sviluppo sostenibile del WHO, come anche la salute in generale. L'approccio metodologico includerà la generazione, la raccolta e l'analisi di un largo set di dati in vivo ed in vitro, comprendendo dati ambientali, molecolari, endocrini, farmacocinetici, genotossici ed epigenetici legati alle funzioni riproduttive. Verranno utilizzati dati epigenetici, clinici e di esposizione, specialmente a distruttori endocrini, con un focus particolare sulle gonadotropine. Deliverables: Algoritmo predittivo per strategie di prevenzione della riproduzione umana e identificazione di nuovi biomarcatori.

## **WP4: Studio di casi per strategie di prevenzione personalizzate**

### **Task 4.1: Valutazione dei fattori di rischio per le patologie cardiovascolari (CVD)**

Usando dati dallo studio di popolazione italiano del progetto CUORE (Health Examination Surveys (OEC 1998-202, OEC/HES 2028-2012, HES 2018-19) e della corte MATISS, testeremo il miglioramento della personalizzazione della valutazione del rischio elaborando funzioni di rischio CVD specifico per sesso o fasce di età; studio dell'utilità di differenti metodi di analisi per elaborare funzioni di valutazioni del rischio CVD individuale (modelli di hazard proporzionale di Cox, tecniche di Intelligenza Artificiale, come i metodi di apprendimento automatico). Deliverables: funzioni aggiornate e strumenti per la valutazione del rischio CVD individuale, specifico nel caso sia per sesso che per fasce di età.

#### **Task 4.2: Modificatori dell'effetto, epigenetica e interazione gene-ambiente: confronto tra diverse coorti da una prospettiva di genere**

Valuteremo come le esposizioni ambientali influiscono sulle conseguenze sulla salute/malattia a lungo termine per il bambino tramite meccanismi epigenetici, al fine di prevedere il rischio associato agli stress precoci nella vita e comprendere se i fattori alimentari possano potenzialmente contrastare i difetti epigenetici indotti dall'inquinamento ambientale, utilizzando dati e campioni da diverse coorti italiane di donne in età fertile e di coppie madre-figlio. Deliverable: modelli predittivi per il rischio associato agli stress precoci nella vita.

#### **Task 4.3: Approccio multi-omico e intelligenza artificiale nelle malattie rare: implementazione degli strumenti per un percorso diagnostico innovativo e per una medicina di precisione nelle malattie fibrotiche**

La fibrosi è lo stadio finale di numerose malattie croniche rare caratterizzate da infiammazione immuno-mediata come la Sclerosi Sistemica (SSc), la Fibrosi Polmonare Idiopatica (IPF), la malattia cronica da trapianto contro l'ospite (cGHVD). La capacità di predire l'esordio e la progressione di malattia è un'esigenza medica non ancora soddisfatta.

i) Biopsie liquide (sangue, liquido di lavaggio bronco-alveolare) e solide (cute, polmone, cuore, fegato) di pazienti afferenti alla nostra rete multicentrica saranno trasferiti alla nostra biobanca, collegata ad un centro di analisi su singola cellula ([www.marchebiobank.it](http://www.marchebiobank.it)), per l'ottenimento di informazioni "omiche" su singola cellula. ii) i dati clinici di ciascun paziente insieme a quelli ottenuti mediante analisi su singola cellula, analisi di laboratorio convenzionali e di imaging saranno trasferiti ad un centro di intelligenza artificiale (Unibo, spoke 2, task 2.2) per essere integrati in un nuovo modello di algoritmo che permetta la stratificazione del rischio di sviluppare fibrosi o di progredire verso forme più severe. Deliverables: una nuova biobanca di campioni di malattie fibrotiche; un nuovo algoritmo che permetta la stratificazione dei pazienti (ad uno stadio iniziale o molto precoce di malattia) in base al rischio di sviluppare una fibrosi progressiva.

#### **Task 4.4: Ricerca "Omica" nell'epidemiologia e nella medicina preventiva**

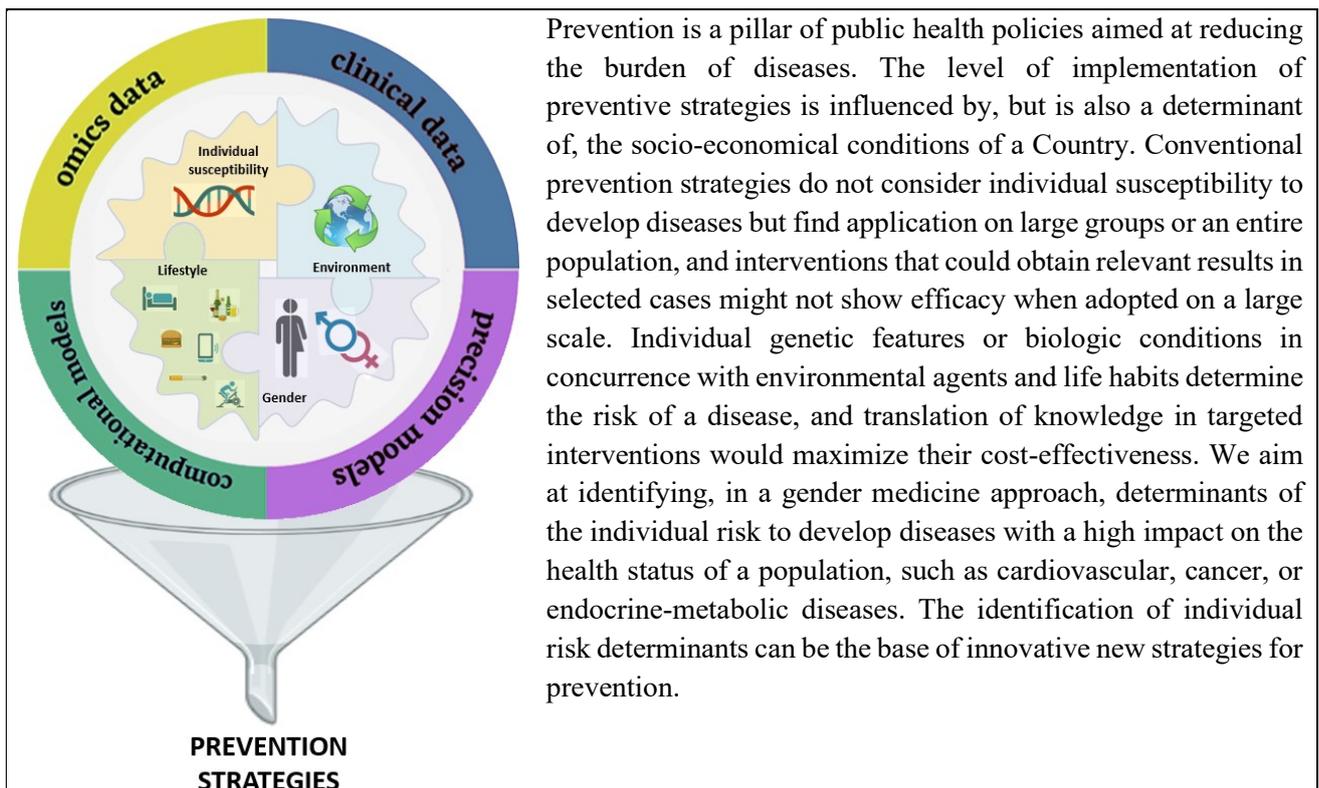
Fornire una comprensione più approfondita della relazione di biomarcatori "omici" su più livelli in corti disponibili. Combinando dati dettagliati su stili di vita stabiliti, fattori di rischio e caratteristiche metaboliche, dati di esposoma e "omica" (metabolomica, epigenomica e genomica), recentemente generati e resi disponibili dal grande consorzio internazionale come biobanca UK, Morgam/BioMarCARE and Nevada Health Genome, algoritmi di medicina di precisione saranno definiti al fine di predire il rischio di gravi malattie croniche. Verranno impiegati metodi di apprendimento automatico per la selezione delle caratteristiche (variabili), la stima delle dimensioni degli effetti e dell'impatto predittivo condizionato al fine di costruire un super "apprendista". Deliverables: Identificazione di markers predisponenti o protettivi, algoritmi personalizzati per la predizione delle patologie croniche.

## SPOKE - 7

### PREVENTION STRATEGIES

Integrated and gender medicine approaches for prevention strategies based on environmental, lifestyle and clinical biometric data

#### GENERAL OBJECTIVE:



#### WP 1: Translation of precision medicine tools into novel strategies for cancer prevention

##### Task 1.1: Predicting cancer development in chronic inflammatory conditions of the digestive system: molecular profiling, individual risk prediction and potential prevention strategies

**1.1.1:** Specific animal models of chronic liver diseases and in vitro facilities will be used to identify markers of liver injury progression and cancer development. Data will be compared with clinical prognostic features of disease progression from local, national and international databases. Specific environmental factors and lifestyle habits will be evaluated as non-transmissible aetiological parameters able to modify morbidity and mortality. A platform for multigenic characterization will be developed, together with tools in support of clinical decision-making. **Deliverable:** molecular profile of patients at risk of developing primary liver cancers

**1.1.2:** Studying the role of bacterial toxins in colorectal tumor- and carcino-genesis, to define genetic profiles associated with increased risk and actionable mechanisms, with the final goal to explore preventive strategies based on dietary or microbioma-modifying interventions. **Deliverable:** Identification of immune, metabolic, epigenetic mechanisms modulated by bacterial toxins.

### Task 1.2: Weighting the exposome: a model for disease prevention

A BioBank for biological samples of former asbestos exposed (ex-EXP) subjects, patients with malignant pleural mesothelioma (MPM) and healthy subjects has been set up. Circulating biomarkers of different origin linked to asbestos exposure, such as asbestos-related miRNAs, autophagic biomarkers of mesothelial transformation ATG5 and HMGB1, and soluble mesothelin related proteins (SMRPs) will be investigated for MPM prediction and follow-up. **Deliverables:** i) an algorithm for prevention and early detection of the asbestos-related cancer diseases; ii) molecular signatures for asbestos-related cancer diseases and follow-up mortality of MPM.

### Task 1.3: Accelerating the development of innovative therapies for solid tumors by advanced preclinical evidence and 3D systems

Microfluidic devices (immune-on-a-chip) generated with the patient's primary cells to analyze the interactions of specific immune cell populations with the host's pathological tissues, and to test the activity of immunotherapies or the effect of conventional therapies on the immune system, in a patient-specific way, facilitating the pathways of personalized medicine. **Deliverable:** Identification of key immune signals driving the effectiveness of cancer therapies.

### Task 1.4: Head and neck oncological predictive medicine by artificial intelligence omic analysis

Identifying genomic and molecular landscape that underlies the phenotypic diversity in HNSCC liquid biopsy using single-cell transcriptomic analysis and next generation sequencing and cells, correlating these findings to specific phenotypes by Artificial Intelligence approaches. **Deliverables:** genomic, transcriptomic, and microbiome biomarkers for predictive diagnostics, prognosis, and drug development in HNSCC patients.

### Task 1.5: Cancer risk prediction in the Italian population

Implementing a novel and comprehensive method for cancer risk prediction in the Italian population, through the outline of comprehensive cancer gene panels (by NGS) as well as the selection of common SNPs to be genotyped for the definition of polygenic risk scores (PRSs) for most frequent cancer types. Algorithms for the evaluation of cancer specific PRSs and tools for the integrative analysis of high/moderate/ low penetrance genetics and environmental exposure will be developed in collaboration with the company "Allelica". **Deliverables:** To provide effective models for the stratification of the Italian population for cancer risk estimates

## WP 2: Novel personalized strategies for prevention of cardiovascular diseases

### Task 2.1: Prevention of Arrhythmogenic Cardiomyopathy (ACM)

Arrhythmogenic Cardiomyopathy (ACM) is a rare, genetic, inherited heart muscle disease that is a main cause of sudden cardiac death (SCD), especially in young adults and athletes and that is frequently misdiagnosed. With laboratory test, ECG recordings, cardiac MRI, an accurate electroanatomical voltage mapping, an endomyocardial biopsy, and genetics, we hypothesize to reduce the portion of misdiagnosis and to better identify phenocopies. **Deliverable:** a diagnostic pathway capable of recognizing ACM patients from pathological phenocopies and a multiparametric score that allows the identification of patients with high arrhythmic risk, using clinical, imaging, electrophysiological, histological and genetic data.

### **Task 2.2: Risk factors assessment for chronic heart failure**

To develop and validate a suitable model for monitoring the density of the  $\beta$ 1AR as a possible predictive marker of heart failure (HF) development using monocytes of the peripheral blood isolated from male and female healthy subjects. **Deliverable:** identification of possible predictive markers of heart HF.

### **Task 2.3: A novel female-specific risk score predictive of atherosclerosis in women for personalized prevention strategies**

Cardiovascular (CV) injury is associated with a significant proportion of deaths in women. Women are exposed to specific sex related CV risk factors including menopause and depression and also the traditional risk factors act differently in women compared to men. We will study three different cohorts from pre- to postmenopause. Machine Learning methods to integrate molecular, phenotypic and lifestyle data will be used generate a novel female-specific risk score predictive of atherosclerosis in women. **Deliverables:** sex-specific cardiovascular risk score ready and validated

### **Task 2.4: Novel early predictors of infections of cardiovascular implantable electronic devices**

The use of cardiovascular implantable electronic devices (CIEDs) is widespread and has been shown to improve patients' quality of life and long-term survival. We propose a deep characterization of patients with suspect CIED infection in order i) to identify the risk factors for systemic vs. local infection; ii) to identify the risk factors for CIED infection relapse or adverse outcomes; iii) to provide novel early predictors of CIED infection. **Deliverables:** a novel stratification tool to predict the risk of CIED re-infection after hardware removal and create a novel algorithm for management of CIED patients after hardware removal.

## **WP 3: Integrating old risk factors and novel predictive models for the prevention of Metabolic and Endocrine Diseases**

### **Task 3.1: Risk factors, lifestyle and new biomarkers in obesity and related diseases**

To study the mechanisms and specific effects of nutrients and metabolic disrupting compounds (MDCs) on adipose tissue and human iPSC-derived organoids in human cohorts taking into account sex, age, health status, dietary intake and cross-talk with other organs, including gut and brain. Validated murine models will be used to verify gut-brain-axis signaling and inflammatory processes involved in metabolic diseases regulation and mood disorders. **Deliverables:** Characterization of specific at-risk groups for metabolic diseases onset related to nutrients and/or MDC effects

### **Task 3.2: Diasmoke - Randomised controlled trial evaluating changes in cardiovascular risk in type 2 diabetic patients who switch to combustion-free nicotine delivery systems.**

Large randomized switching trial investigating changes in CV risk factors and metabolic parameters in diabetic smokers who switch to combustion free nicotine delivery systems - CF NDS (e.g. e-cigarettes, heated tobacco products) compared to continuing smoking. **Deliverables:** identification of CV risk factors changes in diabetic patients

### **Task 3.3: Targeting reproductive functions for novel prevention strategies**

Identifying novel prevention and early treatment strategies for reproduction-related diseases, to ensure sexual and reproductive health, one of the Sustainable Developmental Goals of WHO, as well as health in general.

The methodological approach will include generation, collection and analysis of very large sets of in vivo and in vitro data, covering environmental, molecular, endocrine, pharmacogenetic, genotoxic and epigenetic data around reproductive functions. Epigenetics, clinical and exposure data will be used, especially to endocrine disruptors, for, with a particular focus on gonadotropins. **Deliverables:** Predictive algorithm for human reproductive prevention strategies and novel biomarkers identification.

#### WP 4: Case studies for personalized prevention strategies

##### Task 4.1: Risk factors assessment for Cardio Vascular Diseases (CVD)

Using data from Italian population-based studies of the Progetto CUORE (Health Examination Surveys (OEC 1998-202, OEC/HES 2028-2012, HES 2018-19) and MATISS cohort, we will test the improvement of the risk assessment personalization elaborating sex specific or age-group specific CVD risk functions; study the utility of different analysis methods for elaborating individual CVD risk assessment functions (Cox-proportional hazard models, Artificial Intelligence techniques, such as Machine Learning). **Deliverables:** updated functions and tools for the individual CVD risk assessment, eventually specific for both sexes and for age-groups.

##### Task 4.2: Effect Modifiers, epigenetics and gene-environment interaction: comparison among different cohorts in a gender perspective

We will evaluate how environmental exposures impacts on later-life health/disease consequences for the child via epigenetic mechanisms, to predict the risk associated to early-life stressors and to understand whether dietary factors could potentially counteract environmental pollution-induced epigenomic defects, using data and samples from distinct Italian cohorts of women of childbearing age and of mother-child pairs. **Deliverable:** predictive models for the risk associated to early-life stressors.

##### Task 4.3: Multi-omics and AI approach in rare diseases: implementing an innovative diagnostic pathway and precision medicine tool for fibrotic diseases

Fibrosis is the late stage of many chronic, rare diseases characterized by immune-mediated inflammation such as Systemic Sclerosis (SSc), Idiopathic Pulmonary Fibrosis (IPF), chronic Graft Versus Host Disease (cGVHD). Predicting fibrosis onset and progression is an unmet medical need.

i) Liquid (blood, bronchoalveolar lavage) and tissue (skin, lung, heart, liver) biopsies from patients of our multicenter network will be transferred to our biobank connected to a single cell analysis facility ([www.marchebiobank.it](http://www.marchebiobank.it)) for extraction of single cell omics information; ii) Novel single cell data, conventional laboratory data, imaging data and clinical data from each patient will be transferred to the AI facility (Unibo, spoke 2, task 2.2) for integration into new algorithm models enabling stratification of risk of developing fibrosis and progression to severe forms of fibrosis. **Deliverables:** a novel biobank of fibrotic diseases' samples; a novel algorithm enabling stratification of patients (at very early or early disease steps) based on the risk of developing progressive fibrosis.

##### Task 4.4: Omics Research in Epidemiology and Preventive Medicine

To provide in-depth understanding of the relationship of multilevel omics biomarkers measured in available cohorts. Combining detailed data on established life-styles, risk factors and metabolic features, exposome and "omics" (metabolomic, epigenomic and genomic) data, (newly generated and made available by large international consortia as UK biobank, Morgam/BioMarCARE and Nevada Health Genome, precision



medicine algorithms will be defined to predict the risk of the major chronic disease. Machine learning methods will be employed for feature (variable) selection, estimation of effect sizes and of conditional predictive impact to construct a super learner. *Deliverables:* Identification of predisposing or protecting markers, personalized algorithms for chronic disease prediction.