Blood cell DNA methylation biomarkers in preclinical malignant pleural mesothelioma: the EPIC prospective cohort.

Alessandra Allione[1], Clara Viberti[1], Ilaria Cotellessa[1], Chiara Catalano[1], Elisabetta Casalone[1], Giovanni Cugliari[1], Alessia Russo[1], Simonetta Guarrera[2,3], Dario Mirabelli[4,5], Carlotta Sacerdote[6], Marco Gentile[7], Fabian Eichelmann[8,9], Matthias B. Schulze[8,10], Sophia Harlid[11], Anne Kirstine Eriksen[12], Anne Tjønneland[12,13], Martin Andersson[14], Martijn E.T. Dollé[15], Heleen Van Puyvelde[16], Elisabete Weiderpass[16], Miguel Rodriguez-Barranco[17,18], Antonio Agudo[19], Alicia K Heath[20], María-Dolores Chirlaque[18,21], Thérèse Truong[22], Dzevka Dragic[22,23], Gianluca Severi[22,24], Sabina Sieri[25], Torkjel M Sandanger[26], Eva Ardanaz[18,27], Paolo Vineis[28], Giuseppe Matullo[1,5,29,*].

[1] Department of Medical Sciences, University of Turin, Via Santena 19, 10126 Turin, Italy

[2] IIGM - Italian Institute for Genomic Medicine, c/o IRCCS, Candiolo, Turin, Italy

[3] Candiolo Cancer Institute, FPO - IRCCS, Candiolo, Italy.

[4] Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin, 10126 Turin, Italy.

[5] Interdepartmental Center for Studies on Asbestos and Other Toxic Particulates "G. Scansetti", University of Turin, 10126 Turin, Italy

[6] Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital and Center for Cancer Prevention (CPO), 10126 Turin, Italy

[7] A.O.U. Federico II, Naples , Italy

[8] German Institute of Human Nutrition Potsdam-Rehbruecke, Dept. of Molecular Epidemiology, 14558 Nuthetal, Germany;

[9] German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany

[10] University of Potsdam, Institute of Nutritional Science, 14558 Nuthetal, Germany

[11] Department of Radiation Sciences, Umeå University, SE-901 87 Umeå, Sweden

[12] Danish Cancer Society Research Center, Diet, Genes and Environment, Strandboulevarden 49, 2100 Copenhagen, Denmark
[13] Department of Public Health, University of Copenhagen, Copenhagen, DK
[14] Department of Public Health and Clinical Medicine, Sustainable Health, Umeå University, S-90185 Umeå, Sweden
[15] Centre for Health Protection National Institute for Public Health and the Environment, 3720 BA Bilthoven, the Netherlands
[16] International Agency for Research on Cancer, World Health Organisation, Lyon, France
[17] Escuela Andaluza de Salud Pública (EASP), 18011 Granada, Spain; Instituto de Investigación Biosanitaria ibs.GRANADA, 18012 Granada, Spain
[18] CIBER in Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain.
[19] Unit of Nutrition and Cancer, Catalan Institute of Oncology - ICO, L'Hospitalet de Llobregat, Spain; Nutrition and Cancer Group; Epidemiology, Public Health, Cancer Prevention and Palliative Care Program; Bellvitge Biomedical Research Institute - IDIBELL, L'Hospitalet de Llobregat, Spain.
[20] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom
[21] Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia University, Murcia, Spain
[22] Université Paris-Saclay, UVSQ, Inserm, CESP U1018, «Exposome, Heredity, Cancer and Health» team, Gustave Roussy, Villejuif, France
[23] Centre de recherche sur le cancer de l'Université Laval, Département de Médecine Sociale et Préventive, Faculté de Médecine, Québec, Canada; Centre de recherche du CHU de Québec-Université Laval, Axe Oncologie, Québec, Canada
[24] Department of Statistics, Computer Science and Applications «G. Parenti» (DISIA), University of Florence, Italy
[25] Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano Via Venezian, 1, 20133 Milan – Italy
[26] Department of Community Medicine, UiT The Arctic University of Norway,

Norway
[27] Navarra Public Health Institute, Pamplona, 31003, Spain; IdiSNA, Navarra Institute for Health Research, Pamplona, 31008, Spain
[28] MRC Centre for Environment and Health, Imperial College London, London, United Kingdom
[29] Medical Genetics Unit, AOU Città della Salute e della Scienza, 10126 Turin, Italy

***Corresponding author**: Giuseppe Matullo, Department of Medical Sciences, University of Turin, Turin, Italy. Email: giuseppe.matullo@unito.it.

Keywords: mesothelioma; DNA methylation; prospective nested case-control study; cancer biomarkers

Abbreviations: AUC, area under curve; DNAm, DNA methylation; EPIC, European Prospective Investigation into Cancer and nutrition; EWAS, Epigenome Wide Association Study; miRNA, microRNA; MM, malignant mesothelioma; MPM, malignant pleural mesothelioma; ROC, Receiver operating characteristic.

Novelty & Impact Statements

The aim of this study was to characterise blood biomarkers able to identify asbestos-exposed individuals before MPM symptoms. DNAm changes as non-invasive biomarkers in pre-diagnostic blood samples of MPM cases were investigated for the first time. We identified a signature of nine DNAm biomarkers which can improve the identification of asbestos-exposed individuals at higher MPM risk in order to adopt more intensive monitoring for early disease identification and treatment.

Abstract

Malignant pleural mesothelioma (MPM) is a rare and aggressive cancer mainly caused by asbestos exposure. Specific and sensitive non-invasive biomarkers may facilitate and enhance screening programs for the early detection of cancer.

We investigated DNA methylation (DNAm) profiles in MPM pre-diagnostic blood samples in a case–control study nested in the European Prospective Investigation into Cancer and nutrition (EPIC) cohort, aiming to characterise DNAm biomarkers associated with MPM. From the EPIC cohort, we included samples from 135 participants who developed MPM during 20 years of follow up and from 135 matched, cancer-free, controls. For the discovery phase we selected EPIC participants who developed MPM within five years from enrolment (n=36) with matched controls.

We identified nine differentially methylated CpGs, selected by 10-fold cross-validation and correlation analyses: cg25755428 (*MRI1*), cg20389709 (*KLF11*), cg23870316, cg13862711 (*LHX6*), cg06417478 (*HOOK2*), cg00667948, cg01879420 (*AMD1*), cg25317025 (*RPL17*) and cg06205333 (*RAP1A*). Receiver operating characteristic (ROC) analysis showed that the model including baseline characteristics (age, sex, PC1wbc) along with the nine MPM-related CpGs has a better predictive value for MPM occurrence than the baseline model alone, maintaining some performance also at more than five years before diagnosis [AUC (area under the curve) < 5 years=0.89; AUC 5-10 years=0.80; AUC >10 years=0.75; baseline AUC range=0.63-0.67)].

DNAm changes as non-invasive biomarkers in pre-diagnostic blood samples of MPM cases were investigated for the first time. Their application can improve the identification of asbestos-exposed individuals at higher MPM risk in order to possibly adopt more intensive monitoring for early disease identification.

Introduction

Malignant mesothelioma (MM) is an aggressive cancer of the serous membranes with an increasing incidence worldwide. MM shows a latency period for up to 40 years and its prognosis is poor, with a median survival of about 12 months from diagnosis [1]. At the time of clinical diagnosis, disease is usually unresectable and chemotherapy only vaguely improves prognosis, compared to best supportive care. Currently, the diagnosis of malignant pleural mesothelioma (MPM) is particularly challenging, with no established tissue or soluble biomarkers in clinical practice. The diagnosis of mesothelioma at early stages might be a promising opportunity to improve prognosis over time. Today, most of the MPM studies are focused on biomarkers research, including genetic and epigenetic ones [2,3].

Exposure to asbestos fibres is a major risk factor for MPM, lung cancer, and other non-neoplastic conditions, such as asbestosis and pleural plaques [4]. However, several studies have shown that polymorphisms in genes involved in xenobiotic and oxidative metabolism or in DNA repair processes may modify individual susceptibility to disease after exposure to asbestos [5]. The risk of developing MPM is slightly increased among people exposed to asbestos with a positive history of familial cancers [6].

Several blood-based potential MPM biomarkers have been reported recently in retrospective studies, including peripheral blood DNA methylation (DNAm) variation and microRNAs (miRNAs) [7]: the finding of asbestos fibres in extrapulmonary tissues, translocated through lymphatic and blood flows and

causing continuous exposure of blood cells to asbestos, supports the search for epigenetic changes in blood cell DNA [8].

Epigenetic mechanisms may modulate gene expression without altering the DNA sequence itself. DNAm is a crucial type of epigenetic modification of DNA, by which a methyl group covalently bonds to the C5 position of cytosine in 5'-C-phosphate-G-3' (CpG) dinucleotides; ultimately DNAm regulates gene transcription activity and may modulate key biological functions [9].

Systematic studies of genome-wide DNAm profiles at single CpG level associated with MPM in prospective cohorts are lacking. Most studies so far used biological samples collected at or after diagnosis [10], that may limit application for early disease detection of the identified biomarkers due to reverse causality.

We previously identified DNAm biomarkers in blood taken at MPM diagnosis able to discriminate between MPM and non-MPM asbestos-exposed individuals, assuming that DNAm profile of blood cells may contribute to identify early changes associated with MPM development [11]. More recently, we identified one single-CpG signal in *FKBP5* gene associated with asbestos-exposure as a biomarker of MPM and MPM survival [12,13].

In the current study we investigated, for the first time, the relationship between DNA methylation profiles in pre-diagnostic blood samples and MPM, in a case–control study nested in the prospective EPIC (European Prospective Investigation into Cancer and nutrition) cohort [14].

Our aim was to identify predictive biomarkers of MPM development in asbestos-exposed individuals by DNAm analysis in blood samples.

Materials and Methods

Study design and population

EPIC is a multi-centre cohort study, conducted in 23 centres across 10 European countries, aiming at investigating the aetiological role of biological, lifestyle, and environmental factors in cancer and other chronic diseases. Overall, 521,468 healthy participants were enrolled in EPIC, and followed up on an ongoing basis [14]. In the frame of EPIC cohort, 135 participants developed MPM (EPIC-Meso) after a mean follow-up of 8.3 years (range 0.5-18.8 years). The disease was classified by ICD-10 diagnosis codes (ICD, International Statistical Classification of Diseases, Injuries and Causes of Death) as C38.4 (malignant neoplasm of pleura) in all participants. Cancer endpoint data is based on the latest round of follow-up received from the EPIC centers and centralized at IARC between 2014 and 2016.

Pre-diagnostic DNAm profiles of their blood samples, taken at enrolment, were compared to those of matched controls with no oncological diagnosis.

As one case did not pass the data quality control, the final dataset included 134 cases and 134 matched controls (dataset "MESO_ALL", Table I). Controls were 1:1 matched to the cases by sex, age at enrolment (±1.5 years), study centre, and asbestos-exposure (see "Asbestos-exposure assessment" section).

Cases were divided into 3 subgroups (Table I, MESO_ALL Cases) based on the time between blood sample collection and MPM diagnosis: Group 1, n=36 participants who developed MPM within five years; Group 2, n=40 participants who developed MPM from five to 10 years from the recruitment; and Group 3,

n=58 participants who were diagnosed with MPM more than 10 years after recruitment. Group 1 cases and matched controls (MESO_5YRS) were further analysed to identify differential methylation, and its trend with respect to the time before MPM diagnosis (Table I, MESO_5YRS Cases).

Asbestos-exposure assessment

Occupational information was available from the baseline EPIC questionnaire. It included the occupation at enrolment and data on ever working up to the time of enrolment in 52 at-risk occupations. No information was available on duration of employment and time of first employment. A semi-quantitative job-exposure matrix (JEM) was developed by expert epidemiologists as previously described [15], assigning an "exposure probability" and an "exposure intensity" score to each of the 52 EPIC occupational categories. Both scores were expressed over a numerical scale with four levels: 0 "no exposure", 1 "low", 2 "intermediate", 3 "high" and multiplied to generate an Exposure Index (EI). Various exposure metrics have been derived from the EI, as reported in Supplementary Table S1. For the present study, we used a categorical variable with 3 levels defined as "no exposure" (EI=0), "low exposure" (EI= from 1 to 3), and "high exposure" ($EI \geq 4$).

DNA methylation

DNAm levels were measured in DNA from buffy coat collected at enrolment, using the Infinium Methylation EPIC Bead Chip (>850,000 methylation sites, Illumina, San Diego, California). Laboratory methods for DNA extraction,

BeadChip processing, methylation levels measurement, and data quality controls were previously described [11].

DNA methylation bioinformatics analyses

The average methylation value at each CpG locus, i.e. average "beta (β) value" ranging from 0 to 1, was computed as the ratio of the intensity of the methylated signal over the total signal (unmethylated + methylated).

Data were processed using ChAMP Bioconductor package (v2.22) filtering out the BeadChip probes with a detection P value>0.01 and with <3 beads in at least 5% of samples per probe; non-CpG sites SNP-related and CpGs located in chromosome X and Y were excluded. The remaining signals were quality controlled, normalized and batch effect corrected using ComBat algorithm.

After quality controls 725,050 CpGs were used in the following analyses.

DNA methylation regional differences were calculated by ChAMP approach (champ DMR function).

White blood cells estimate

In order to evaluate differences in white blood cells (WBCs) proportion between cases and controls, we estimated WBC subtype percentages from genome-wide methylation data[16]. For each individual, we extracted WBC subtype percentages, estimated based on genome wide methylation data. We estimated the frequencies of B cells, CD8+ and CD4+ T cells, natural killer cells, granulocytes and monocytes using methylation profiles, both in MESO_ALL and MESO_5YRS datasets. WBCs differences between cases and controls were tested with wilcox.test R function.

Statistical analyses

All statistical analyses were conducted using the open source software Rv4.1.0. Epigenome-wide differential methylation on MESO_ALL dataset (134 cases and 134 controls, Table I) was tested by a multiple regression model (R function *glm*).

Cross-validation analysis was conducted on the subset of the original group consisting of the 36 participants who developed MPM within 5 years from the blood sample collection and 36 matched controls (dataset "MESO_5YRS"), in order to improve the estimate of the mean model performance. Cross validation was carried out mixing and splitting the dataset randomly, summarizing the goodness of the model using the sample of model evaluation scores. Each run consisted of an Epigenome Wide Association Study (EWAS) analysis done on 80% of the samples (n=58), with the option of maintaining comparable proportions of matched cases and controls in the two groups; we conducted a total of ten cycles. As at False Discovery Rate (FDR)-level we did not obtain any statistically significant differentially methylated signal between cases and controls (possibly due to the small sample size population), we focused on those signals that showed an |effect size| > 10% and a nominal p $\leq$0.05 in at least 8 out of 10 runs.

All the analyses, including trend test, were adjusted for age at blood collection, sex and WBC's first principal component (PC1wbc; R prcomp function), which takes into account 66% of the variance. WBC subtypes were estimated as previously described [17].

A Pearson correlation test was performed to evaluate the association between two CpG sites, in order to select only one of the two correlated signals (selecting correlating values of |rho| > 0.7, p≤0.05). For each correlated couple, we selected the CpG signal with the higher effect size.

Multiple regression models were done with glm R function, correlations with cor.test R function, ROC analyses with ROCR and pROC libraries.

Results

A flow diagram of participants and relative analyses performed in the present study is represented in Supplementary Figure S1.

Differential methylation analysis

The baseline characteristics of the study participants are shown in Table I.

Leukocyte DNAm was used to quantify different leukocyte sub proportions applying an implemented deconvolution pipeline for high-resolution immune profiling [16]. In MESO_ALL cases we observed a slight difference in neutrophils, that were higher than controls (Wilcoxon test p=0.04, Supplementary Table S2) and a weak decrease of CD4+ memory cells (Wilcoxon test p=0.02, Supplementary Table S2). In MESO_5YRS we did not observe WBC differences between cases and controls (Supplementary Table S3). However, although we expected a larger deviation of WBCs close to diagnosis, in both participant groups we did non observe any trend of WBCs numbers with the time to diagnosis (data not shown).

We first focused the DNAm analysis on 36 participants who developed MPM within five years from recruitment, versus 36 controls matched by sex, age,

study centre and asbestos-exposure level (MESO_5YRS), under the hypothesis that the most significant changes in blood should happen closer to MPM diagnosis, due to the already ongoing carcinogenic process.

In the cross-validation analysis we identified 22 differentially methylated CpGs with nominal $p \leq 0.05$ and |effect size| > 10% (Table II). Unsupervised hierarchical clustering heatmap of DNA methylation of the 22 CpGs is shown in Supplementary Figure S2.

We observed a very high correlation between DNA methylation levels of CpGs located in the same gene (Supplementary Table S4), allowing us to select only one CpG for each gene, with a final list of nine significant CpGs (highlighted in bold in Table II).

Regional differences in MESO_5YRS cases and controls were analysed with no statistically significant results (Supplementary Table S5). Looking at DMRs with a mean of the estimate $> |0.5|$, we confirmed DMRs in genes already identified in Table II: LHX6 (16 CpGs), HOOK2 (4 CpGs), RAP1A (3 CpGs).

The association between methylation of 22 CpGs listed in Table II and asbestos-exposure was evaluated both in cases and in controls. To do this, we performed a Trend test, adjusted for age, sex and PC1wbc in MESO_5YRS individuals (26 cases and 26 controls with exposure assessment). However, we did not observe any statistically significant results (Supplementary Table S6).

EWAS analysis performed on the whole population of 268 participants (MESO_ALL) showed three signals with nominal $p \leq 0.05$ and |effect size| > 10%: cg04131969 (*MYADML* gene, effect size=0.11, p=0.004) was hypermethylated in cases compared with controls, while cg17939448 (*FAM47E*

gene, effect size=-0.10, p=0.001) and cg01201512 (*NINJ2* gene, effect size=-0.10, p=0.006) were hypomethylated.

Receiver operating characteristic (ROC) curves of DNA methylation signals

In order to evaluate improvement in discrimination between cases and controls when including single CpG methylation levels in the analysis (nine CpGs obtained by cross-validation analysis and selected after correlation analysis), we compared two ROC analyses: model 0 includes age, sex, PC1wbc and model 1 includes age, sex, PC1wbc and single CpG methylation levels.

ROC analyses (one for each of the 9 CpGs identified in MESO_5YRS) were performed on a larger population of 107 participants, including the 71 tumour-free asbestos-exposed (asbestos-exposure = 1 or 2) participants from MESO_ALL and 36 participants who developed MPM within five years from recruitment from MESO_5YRS (Table III). One DNAm signal was statistically significant: cg23870316 (AUC model 0= 0.66; AUC model 1= 0.75; De-Long's p=0.02). However, DNA methylation level of cg23870316 is not able to discriminate cases from controls in participants recruited more than five years before MPM diagnosis (Supplementary Table S7).

In order to investigate a potential extension of results to longer time windows, we also carried out ROC analysis with all the nine identified CpGs, to discriminate asbestos-exposed controls with no diagnosis of MPM during follow up (n=71, asbestos-exposure = 1 or 2) from participants who developed MPM within five years after enrolment (n=36, Fig. 1a), within 5-10 years (n=40,

Fig. 1b), and over 10 years (n=58, Fig. 1c). We performed ROC analysis including model 0: age, sex, PC1wbc; and model 1: age, sex, PC1wbc, 9 CpGs methylation levels. De-Long's test showed a statistically significant discrimination between the two groups in model 1 compared to model 0 in all the subgroups: within 5 years from MPM (AUC model 0=0.66, AUC model 1=0.89, De-Long's $p=6.4x10^{-5}$), from 5 to 10 years (AUC model 0=0.67, AUC model 1=0.80, De-Long's p=0.018), and > 10 years (AUC model 0=0.63, AUC model 1=0.75, De-Long's p=0.018), showing a good performance also when the participants were recruited > 10 years before MPM diagnosis.

ROC analysis with the nine CpG was also performed excluding the 20 asbestos-exposed controls belonging to the MESO_5YRS dataset and already used in discovery. Thus, ROC curves were performed including model 0: age, sex, PC1wbc; and model 1: age, sex, PC1wbc, nine CpGs methylation levels on 51 asbestos-exposed controls and 36 MESO_5YRS cases, obtaining similar results (AUC model 0= 0.70; AUC model 1= 0.87; De-Long's p= 0.004). Moreover, to exclude asbestos-exposure effect in case-control discrimination, we performed also the same ROC analysis with model 0 (age, sex, PC1wbc, asbestos-exposure) and model 1 (age, sex, PC1wbc, asbestos-exposure, nine CpGs methylation levels), confirming previous results (AUC model 0= 0.73; AUC model 1= 0.86; De-Long's p= 0.02).

Biomarkers trend with time to diagnosis

We performed a trend test to evaluate the changes of the 9 differentially methylated CpGs in association with the time to diagnosis. MESO_5YRS MPM cases were divided into participants who developed MPM within 1.5 years from

sample collection (n=9), participants who developed MPM after 1.5 and before 3 years (n=13) and participants who developed MPM after 3 and before 5 years (n=14) (MESO_5YRS cases, Tab. I) and the trend test was performed comparing these three groups. We did not observe any statistically significant trend (Supplementary Table S8). On the other hand, three of these signals showed a statistically significant linear trend in MESO_ALL cases divided by time from recruitment to diagnosis as described in Tab. I (Fig. 2): cg01879420 in *AMD1* gene (p=0.006), cg25755428 in *MRI1* gene (p=0.021) and cg23870316 (p=0.001). Nevertheless, EPIC participants who were diagnosed with MPM a longer time from enrolment showed DNAm levels very similar to control participants, as expected (Fig. 2).

DNAm diagnostic biomarkers changes in pre-diagnostic samples

We aimed to assess whether previously reported DNAm biomarkers in MPM patients at diagnosis [11,13] are also informative in DNA blood samples taken up to five years before MPM diagnosis. To do this, we performed a multiple regression model on MESO_5YRS dataset of DNAm biomarkers previously identified as differentially methylated in a case/control study (Supplementary Table S9). Cg01521397 in *TAF4* gene body was statistically significantly hypomethylated in EPIC-Meso who developed MPM within five years from recruitment compared to matched controls (p=0.01). However, any of these CpGs showed statistically significant differences in individuals closer to MPM diagnosis (data not shown) compared to the other participants.

Discussion

In the present study we identified nine DNAm biomarkers in blood pre-clinical samples from MPM patients. The panel of the nine signals was able to discriminate cases from controls with the highest performance within five years before diagnosis, and maintaining some discriminating power up to ten years before MPM symptoms. Generally, the diagnosis of cancer at early stages, when clinical symptoms have not yet occurred, appears to be a promising opportunity to improve therapeutic outcomes. Treatment at early stages together with newly developed therapies could possibly lead to improvement in overall survival of patients with MM [18]. We focused the DNAm data analysis mainly on individuals who were diagnosed with MPM within five years from the blood sample collection, with the purpose to identify early changes of tumour-related DNAm. A few previous studies measured blood biomarkers several years before MPM diagnosis, but comparison with them is hard due to different time from measurements to MPM diagnosis, most of them shorter than those of EPIC [16,19]. Except for mesothelin [19], no other biomarkers have been evaluated in a sufficiently large prospective cohort study with serial pre-diagnostic samples of MM cases so far. Morrè and colleagues, observed that ENOX2 protein transcript variants characteristic of malignant mesothelioma were present in serum 4-10 years in advance of clinical symptoms [20].

Among the 22 differentially methylated signals, we identified five genes in which more than one CpG were differentially methylated: *LHX6*, *MRI1*, *HOOK2*, *AMD1* and *KLF11*.

In particular, *LHX6* gene contained 10 hypermethylated CpGs (Supplementary Fig. S3 shows a gene-based visualization of DM CpGs in LHX6 gene), two of which were located in transcriptionally active gene regions (e.g., TSS200, 1st exon). Methylation-mediated inactivation of *LHX6* (LIM Homeobox domain 6) has been shown in lung [21], head and neck [22], cervical [23] and pancreatic [24] tumours. The mechanism by which *LHX6* acts as a tumour suppressor gene is related to the interference with the Wnt/β-catenin pathway [25], which is severely involved also in MPM progression [26]. Moreover, *LHX6* is a target gene of miR-214, and upregulated *LHX6* gene expression was observed to be related to downregulation of miR-214 in non-small-cell lung cancer [27]. The role of miR-214 has also been evaluated in mesothelioma, and previous studies in human samples showed a down-regulation of miR-214 expression [28].

Although it is already known that LHX6-cg13862711 hypermethylation in tissue is associated with several types of cancer (e.g., in breast, kidney and upper aerodigestive tract) (https://cancer.sanger.ac.uk/cosmic), the epigenetic regulation of *LHX6* in blood cells of pre-clinical samples needs further investigation.

Two hypomethylated signals in cases compared to controls in the *KLF11* 5'UTR/TSS200 region were identified. Krüppel-like factors (KLFs) form a highly conserved family of zinc finger transcription factors, and play important roles in the progression of human malignant tumours, such as breast cancer and colon cancer [29]. Recently, high *KLF11* expression was associated with poor prognosis of glioma [30].

In *MRI1* gene we identified two hypermethylated signals located in the promoter region (TSS1500) in cases compared to controls. *MRI1* codes for an enzyme (methylthioribose-1-phosphate isomerase) that catalyses the interconversion of methylthioribose-1-phosphate (MTR-1-P) into methylthioribulose-1-phosphate (MTRu-1-P). In addition to the catalytic activity, MRI1 promotes cell invasion and signal transmission in response to RhoA activation in cancer cells, and for this reason is also called "Mediator of RhoA-dependent Invasion" (MRDI) [31]. Elevated expression of the encoded protein is associated with metastatic melanoma and this protein promotes melanoma cell invasion independently of its enzymatic activity [31].

In the gene body of *HOOK2* (Hook Microtubule Tethering Protein 2) we identified two hypomethylated CpGs. Several studies focused on differentially methylated regions in *HOOK2*, showing its implication in several pathological conditions, as in diabetes, obesity, polycystic ovary syndrome and cardiometabolic diseases [32-34]. In particular, the deregulation of cg06417478 DNAm was shown as associated with diabetes [34], while DNAm changes in cg23899408 were identified in blood as a prediction marker in liver diseases [35]. Deregulation of both CpGs was associated with colon adenocarcinoma in patients with metabolic syndrome [36].

Other deregulated CpGs were found in *AMD1* (Adenosylmethionine Decarboxylase 1) gene. This gene encodes S-adenosylmethionine decarboxylase 1, an important intermediate enzyme in polyamine biosynthesis. The polyamines spermine, spermidine, and putrescine are highly regulated in cellular proliferation and tumour promotion [37]. Multiple alternatively spliced

transcript variants have been identified, also in association with cancer survival (e.g., non-small cell lung cancer) [38]. Several studies have demonstrated that *AMD1* differential expression is implicated in cancer, such as chronic myeloid leukaemia (CML) [39] and B-cell non-Hodgkin's lymphoma [40].

Finally, two other genes contain significant differentially methylated CpG sites: *RPL17* (cg25317025) and *RAP1A* (cg06205333).

RPL17 encodes a ribosomal protein that is a component of the large 60S subunit, called 60S ribosomal protein L17 [41]. A variety of extra-ribosomal functions were recently recognized for ribosomal proteins, including the regulation of immune signalling, tumorigenesis and cellular development [42]. RPL17 expression has been reported to be associated with breast [43] and liver [44] cancer.

RAP1A deregulated CpG is located in the promoter region (TSS1500). *RAP1A* encodes a small GTPase member of the Ras family. Its encoded protein may be involved in signalling pathways that affect cell proliferation, adhesion, and may play a role in tumour malignancy [45,46]. Inactivation of RAP1 by bisphosphonates treatment has been observed in several pathologies, including MM: growth inhibition of MM cells by zoledronic acid has been shown both *in vitro* and *in vivo*, and the described effect is related to RAP1 unprenylation [47,48].

Besides the seven genes described above, we also identified two deregulated CpGs in non-coding regions (cg23870316 and cg00667948). In this regard, the cg23870316 is the unique DNAm signal which was statistically significant in

discriminating prospective cases from controls. This CpG is located in the 8p23.2 DNA region with scanty information in the literature.

The nine differentially methylated CpGs together were not only able to discriminate cases within five years from MPM diagnosis, but they also partly maintained their discrimination power also for cases who were diagnosed with MPM in a range of 5-10 years or more than 10 years from sample collection. However, the significance (De-Long's test) decreases moving away from the time of diagnosis, showing that the pre-diagnostic potential of these biomarkers is limited to time periods closer to diagnosis.

In summary, our results identified nine differentially methylated CpGs that could be used as MPM biomarkers to identify early changes in pre-clinical individuals (i.e., within five years before diagnosis). Interestingly, the identified CpGs are located on genes already known for their potential role as tumour biomarkers.

Regarding the estimate of WBC subpopulations, we applied a new reference-based deconvolution analysis of peripheral blood DNA methylation data developed by Salas et al., which include memory and naïve cells from cytotoxic and helper T cells and B cells and parse the granulocyte subtypes into neutrophils, eosinophils, and basophils[16]. Although we did not obtain statistically significant results in MESO_5YRS dataset, we observed a statistically significant difference between pre-clinical cases and controls in CD4+ memory T cells and neutrophils in MESO_ALL dataset, indicating a different immunological response in individuals who developed MPM compared to controls. The reduction of estimated CD4+ memory lymphocytes in cases

suggests a weaker adaptive immune system and is compatible with the possible occurrence of functional changes in cellular subpopulations in MPM, while an increase of neutrophils could correlate with the recent interest for their cancer-promoting effects: an elevated neutrophil-to-lymphocyte ratio is considered a prognostic indicator for cancer patients, in particular in mesothelioma, although its value is still debated [49]. However, the WBCs differences that we estimated did not change in relation to the time to diagnosis, and are so minimal that it could need further insights.

We previously observed DNAm biomarkers in MPM patients [11,13]: among them only cg01521397 in *TAF4* gene is differentially hypomethylated in pre-clinical samples (as well as in MPM patients), but with any significant changes closer to MPM diagnosis. The meaning of this results is interesting in the context of biomarkers reverse causality interpretation, discriminating biomarkers related to the presence of MPM from those associated to prediagnostic biological changes. However, if reverse causality accounted for our findings, we would expect the cg01521397 DNAm-MPM association to become stronger closer to the time of diagnosis and this does not seem to be the case.

Furthermore, we are aware of the limitations of our study. First of all the lack of a replication in an independent cohort. Small sample size related problems are especially common in the study of pre-clinical samples of rare diseases. For this reason, cross-validation represents a common and efficient solution when the available data are limited. In addition, when validation with a separate dataset is not feasible, cross-validation allows to use all the data for training and to reuse it for validation. The procedure of splitting data at every run

offers an efficient quality forecast and we were also able to make predictions on all our data. Moreover, each run is independent from the others, avoiding overestimation of the generalization of the model by/through developing a new one with each CV cycle. The cross-validation approach is widely used, representing a very powerful tool: reduces bias, improves the use of the data, avoiding the overfitting problem, especially with small datasets, produces robust and unbiased performance estimates regardless of sample size [50]. However, this limitation needs to be considered interpreting results.

The use of whole blood samples to examine the DNAm levels, which may not directly reflect the status of the target tissue, is another limit of the study. However, the analysis of pre-clinical samples with the main goal to monitor high MPM risk asbestos-exposed individuals should rely on a non-invasive biosample such as blood.

The coverage at disease relevant genes by CpG sites included in the DNAm Infinium Methylation EPIC beadchip array used in the study could be another limitation, as it investigates only 30% of the human methylome. In comparison, whole-genome bisulfite sequencing is able to capture more than 28 million CpGs, but the feasibility remains low for the population-based EWAS due to high cost and large genomic DNA input requirements to compensate for degradation during DNA bisulfite treatment.

The large EPIC cohort is a unique and very valuable resource, as the prospective nature of the study is more suitable to stratify high-risk individuals on the basis of genetic and epigenetic profiles in combination with other biomarkers and clinical risk factors. Moreover, since MPM diagnosis takes place

only in the late stage of the disease, prospective cohort studies may be optimal for investigation of early MPM-related changes in the asymptomatic phase.

The identified DNAm biomarkers could be associated with germline mutations analysis in the follow-up of asbestos-exposed individuals with MPM genetic predisposition. Based on published studies from our group and others, germline mutations in several genes may increase MPM susceptibility even with low levels of asbestos-exposure, and possibly predict the response to standard treatments [51,52]. Moreover, the detection of *BRCA*-mutated or *BAP1*-mutated MPM in a very early phase of tumour development could address a personalized therapy approach with PARP-inhibitors or immune checkpoint inhibitors [53].

This study identified potential blood DNAm changes in pre-clinical MPM individuals. If further replicated in other studies, these signals could represent potential circulating biomarkers, promising for early MPM detection.

Acknowledgements

We thank Dr. Lucas A. Salas (Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA) for his kindly and precious support in WBC estimation.

Author Contributions

The work reported in the paper has been performed by the authors, unless clearly specified in the text

CRediT author statement: Alessandra Allione: conceptualization, methodology, writing – original draft, project administration; Clara Viberti: formal analysis, data curation, writing – original draft; Ilaria Cotellessa: formal analysis, data curation, writing – original draft; Chiara Catalano: investigation; Elisabetta Casalone: conceptualization, methodology, project administration; Giovanni Cugliari: formal analysis, data curation; Alessia Russo: methodology; Simonetta Guarrera: methodology; Dario Mirabelli: data curation; Carlotta Sacerdote: resources, data curation; Marco Gentile: resources, data curation; Fabian Eichelmann: resources, data curation; Matthias B. Schulze: resources, data curation, writing - review & editing; Sophia Harlid: resources, data curation; Anne Kirstine Eriksen: resources, data curation; Anne Tjønneland: resources, data curation; Martin Andersson: resources, data curation; Martijn Dollé: resources, data curation; Heleen Van Puyvelde: resources, data curation; Elisabete Weiderpass: resources, data curation; Miguel Rodriguez-Barranco: resources, data curation; Antonio Agudo: resources, data curation; Alicia K Heath: resources, data curation; María-Dolores Chirlaque: resources, data curation; Thérèse Truong: resources, data curation; Dzevka Dragic: resources, data curation; Gianluca Severi: resources, data curation; Sabina Sieri: resources, data curation; Torkjel M Sandanger: resources, data curation; Eva Ardanaz: resources, data curation; Paolo Vineis: conceptualization; Giuseppe Matullo: conceptualization, supervision, project administration, funding acquisition.

Conflict of interest

The authors declare no financial competing interests.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Data availability statement

Array data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001006432. Further information about EGA can be found on https://ega-archive.org "The European Genome-phenome Archive of human data consented for biomedical research" (http://www.nature.com/ng/journal/v47/n7/full/ng.3312.html). Other data that support the findings of this study are available from the corresponding author upon request.

Ethics Statement

The study was approved by the Institutional Review Board of IARC and the ethics committees in the participating countries. All participants gave their written informed consent to participate in the EPIC study. The present study

was carried out in accordance with the ethical principles of the Declaration of Helsinki.

Funding

The research leading to these results has received funding from AIRC under IG 2018 - ID. 21390 project – P.I. G. Matullo and by Ministero dell'Istruzione, dell'Università e della Ricerca – MIUR project "Dipartimenti di Eccellenza 2018 – 2022" (n° D15D18000410001, to G.M.) to the Department of Medical Sciences, University of Torino.

The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC).

The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer

Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford) (United Kingdom).

References

1. Cakiroglu E, Senturk S. Genomics and Functional Genomics of Malignant Pleural Mesothelioma. *Int J Mol Sci.* Sep 1 2020;21(17).

2. Carbone M, Adusumilli PS, Alexander HR, Jr., et al. Mesothelioma: Scientific clues for prevention, diagnosis, and therapy. *CA Cancer J Clin.* Sep 2019;69(5):402-429.

3. Ferrari L, Carugno M, Mensi C, Pesatori AC. Circulating Epigenetic Biomarkers in Malignant Pleural Mesothelioma: State of the Art and critical Evaluation. *Front Oncol.* 2020;10:445.

4. Fazzo L, Binazzi A, Ferrante D, et al. Burden of Mortality from Asbestos-Related Diseases in Italy. *Int J Environ Res Public Health.* Sep 23 2021;18(19).

5. Neri M, Ugolini D, Dianzani I, et al. Genetic susceptibility to malignant pleural mesothelioma and other asbestos-associated diseases. *Mutat Res.* Jul-Aug 2008;659(1-2):126-136.

6. Ugolini D, Neri M, Ceppi M, et al. Genetic susceptibility to malignant mesothelioma and exposure to asbestos: the influence of the familial factor. *Mutat Res.* Mar-Apr 2008;658(3):162-171.

7. Sinn K, Mosleh B, Hoda MA. Malignant pleural mesothelioma: recent developments. *Curr Opin Oncol.* Jan 2021;33(1):80-86.

8. Fubini B, Otero Arean C. Chemical aspect of the toxicity of inhaled mineral dusts. *The Royal Society of Chemistry.* 1999;28:373-381.

9. Kanherkar RR, Bhatia-Dey N, Csoka AB. Epigenetics across the human lifespan. *Front Cell Dev Biol.* 2014;2:49.

10. Foddis R, Bonotti A, Landi S, Fallahi P, Guglielmi G, Cristaudo A. Biomarkers in the prevention and follow-up of workers exposed to asbestos. *J Thorac Dis.* Jan 2018;10(Suppl 2):S360-S368.

11. Guarrera S, Viberti C, Cugliari G, et al. Peripheral Blood DNA Methylation as Potential Biomarker of Malignant Pleural Mesothelioma in Asbestos-Exposed Subjects. *J Thorac Oncol.* Mar 2019;14(3):527-539.
12. Cugliari G, Catalano C, Guarrera S, et al. DNA Methylation of FKBP5 as Predictor of Overall Survival in Malignant Pleural Mesothelioma. *Cancers (Basel).* Nov 21 2020;12(11).
13. Cugliari G, Allione A, Russo A, et al. New DNA Methylation Signals for Malignant Pleural Mesothelioma Risk Assessment. *Cancers (Basel).* May 27 2021;13(11).
14. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* Dec 2002;5(6B):1113-1124.
15. Pesch B, Gawrych K, Rabstein S, et al. N-acetyltransferase 2 phenotype, occupation, and bladder cancer risk: results from the EPIC cohort. *Cancer Epidemiol Biomarkers Prev.* Nov 2013;22(11):2055-2065.
16. Salas LA, Zhang Z, Koestler DC, et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat Commun.* Feb 9 2022;13(1):761.
17. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* May 8 2012;13:86.
18. Scherpereel A. Malignant pleural mesothelioma: new treatments, new hopes? *Eur Respir J.* Mar 2017;49(3).
19. Filiberti R, Marroni P, Spigno F, et al. Is soluble mesothelin-related protein an upfront predictive marker of pleural mesothelioma? A prospective study on Italian workers exposed to asbestos. *Oncology.* 2014;86(1):33-43.
20. Morre DJ, Hostetler B, Taggart DJ, et al. ENOX2-based early detection (ONCOblot) of asbestos-induced malignant mesothelioma 4-10 years in advance of clinical symptoms. *Clin Proteomics.* 2016;13:2.
21. Liu WB, Jiang X, Han F, et al. LHX6 acts as a novel potential tumour suppressor with epigenetic inactivation in lung cancer. *Cell Death Dis.* Oct 24 2013;4:e882.
22. Estecio MR, Youssef EM, Rahal P, et al. LHX6 is a sensitive methylation marker in head and neck carcinomas. *Oncogene.* Aug 17 2006;25(36):5018-5026.
23. Jung S, Jeong D, Kim J, et al. Epigenetic regulation of the potential tumor suppressor gene, hLHX6.1, in human cervical cancer. *Int J Oncol.* Mar 2011;38(3):859-869.
24. Abudurexiti Y, Gu Z, Chakma K, et al. Methylation-mediated silencing of the LIM homeobox 6 (LHX6) gene promotes cell proliferation in human pancreatic cancer. *Biochem Biophys Res Commun.* Jun 4 2020;526(3):626-632.
25. Yang J, Han F, Liu W, et al. LHX6, An Independent Prognostic Factor, Inhibits Lung Adenocarcinoma Progression through Transcriptional Silencing of beta-catenin. *J Cancer.* 2017;8(13):2561-2574.

26. Fox SA, Richards AK, Kusumah I, et al. Expression profile and function of Wnt signaling mechanisms in malignant mesothelioma cells. *Biochem Biophys Res Commun.* Oct 11 2013;440(1):82-87.
27. Liao J, Lin J, Lin D, et al. Down-regulation of miR-214 reverses erlotinib resistance in non-small-cell lung cancer through up-regulating LHX6 expression. *Sci Rep.* Apr 10 2017;7(1):781.
28. Amatya VJ, Mawas AS, Kushitani K, Mohi El-Din MM, Takeshima Y. Differential microRNA expression profiling of mesothelioma and expression analysis of miR-1 and miR-214 in mesothelioma. *Int J Oncol.* Apr 2016;48(4):1599-1607.
29. Cheng L, Shi L, Dai H. Bioinformatics analysis of potential prognostic biomarkers among Kruppel-like transcription Factors (KLFs) in breast cancer. *Cancer Biomark.* 2019;26(4):411-420.
30. Xi Z, Zhang R, Zhang F, Ma S, Feng T. KLF11 Expression Predicts Poor Prognosis in Glioma Patients. *Int J Gen Med.* 2021;14:2923-2929.
31. Templeton PD, Litman ES, Metzner SI, Ahn NG, Sousa MC. Structure of mediator of RhoA-dependent invasion (MRDI) explains its dual function as a metabolic enzyme and a mediator of cell invasion. *Biochemistry.* Aug 20 2013;52(33):5675-5684.
32. Jacobsen VM, Li S, Wang A, et al. Epigenetic association analysis of clinical sub-phenotypes in patients with polycystic ovary syndrome (PCOS). *Gynecol Endocrinol.* Aug 2019;35(8):691-694.
33. Kraus WE, Muoio DM, Stevens R, et al. Metabolomic Quantitative Trait Loci (mQTL) Mapping Implicates the Ubiquitin Proteasome System in Cardiovascular Disease Pathogenesis. *PLoS Genet.* Nov 2015;11(11):e1005553.
34. Rizzo HE, Escaname EN, Alana NB, et al. Maternal diabetes and obesity influence the fetal epigenome in a largely Hispanic population. *Clin Epigenetics.* Feb 19 2020;12(1):34.
35. Li K, Qin L, Jiang S, et al. The signature of HBV-related liver disease in peripheral blood mononuclear cell DNA methylation. *Clin Epigenetics.* Jun 8 2020;12(1):81.
36. Chitrala KN, Hernandez DG, Nalls MA, et al. Race-specific alterations in DNA methylation among middle-aged African Americans and Whites with metabolic syndrome. *Epigenetics.* May 2020;15(5):462-482.
37. Casero RA, Jr., Murray Stewart T, Pegg AE. Polyamine metabolism and cancer: treatments, challenges and opportunities. *Nat Rev Cancer.* Nov 2018;18(11):681-695.
38. Chen K, Liu H, Liu Z, et al. Genetic variants in RUNX3, AMD1 and MSRA in the methionine metabolic pathway and survival in nonsmall cell lung cancer patients. *Int J Cancer.* Aug 1 2019;145(3):621-631.
39. Sari IN, Yang YG, Wijaya YT, et al. AMD1 is required for the maintenance of leukemic stem cells and promotes chronic myeloid leukemic growth. *Oncogene.* Jan 2021;40(3):603-617.
40. Scuoppo C, Miething C, Lindqvist L, et al. A tumour suppressor network relying on the polyamine-hypusine axis. *Nature.* Jul 12 2012;487(7406):244-248.

41. Wang M, Parshin AV, Shcherbik N, Pestov DG. Reduced expression of the mouse ribosomal protein Rpl17 alters the diversity of mature ribosomes by enhancing production of shortened 5.8S rRNA. *Rna.* Jul 2015;21(7):1240-1248.
42. Zhou X, Liao WJ, Liao JM, Liao P, Lu H. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol.* Apr 2015;7(2):92-104.
43. Yuan F, Wang W, Cheng H. Co-expression network analysis of gene expression profiles of HER2(+) breast cancer-associated brain metastasis. *Oncol Lett.* Dec 2018;16(6):7008-7019.
44. Xing M, Wang X, Kiken RA, He L, Zhang JY. Immunodiagnostic Biomarkers for Hepatocellular Carcinoma (HCC): The First Step in Detection and Treatment. *Int J Mol Sci.* Jun 7 2021;22(11).
45. Li H, Han G, Li X, et al. MAPK-RAP1A Signaling Enriched in Hepatocellular Carcinoma Is Associated With Favorable Tumor-Infiltrating Immune Cells and Clinical Prognosis. *Front Oncol.* 2021;11:649980.
46. Li Q, Xu A, Chu Y, et al. Rap1A promotes esophageal squamous cell carcinoma metastasis through the AKT signaling pathway. *Oncol Rep.* Nov 2019;42(5):1815-1824.
47. Merrell MA, Wakchoure S, Ilvesaro JM, et al. Differential effects of Ca(2+) on bisphosphonate-induced growth inhibition in breast cancer and mesothelioma cells. *Eur J Pharmacol.* Mar 15 2007;559(1):21-31.
48. Okamoto S, Kawamura K, Li Q, et al. Zoledronic acid produces antitumor effects on mesothelioma through apoptosis and S-phase arrest in p53-independent and Ras prenylation-independent manners. *J Thorac Oncol.* May 2012;7(5):873-882.
49. Chen N, Liu S, Huang L, et al. Prognostic significance of neutrophil-to-lymphocyte ratio in patients with malignant pleural mesothelioma: a meta-analysis. *Oncotarget.* Aug 22 2017;8(34):57460-57469.
50. Valente G, Castellanos AL, Hausfeld L, De Martino F, Formisano E. Cross-validation and permutations in MVPA: Validity of permutation strategies and power of cross-validation schemes. *Neuroimage.* Sep 2021;238:118145.
51. Betti M, Aspesi A, Sculco M, Matullo G, Magnani C, Dianzani I. Genetic predisposition for malignant mesothelioma: A concise review. *Mutat Res Rev Mutat Res.* Jul - Sep 2019;781:1-10.
52. Panou V, Gadiraju M, Wolin A, et al. Frequency of Germline Mutations in Cancer Susceptibility Genes in Malignant Mesothelioma. *J Clin Oncol.* Oct 1 2018;36(28):2863-2871.
53. Fennell DA, King A, Mohammed S, et al. Rucaparib in patients with BAP1-deficient or BRCA1-deficient mesothelioma (MiST1): an open-label, single-arm, phase 2a clinical trial. *Lancet Respir Med.* Jun 2021;9(6):593-600.

Figure Legends

Fig. 1. ROC curves including model 0 (dotted line): age, sex, PC1wbc; and model 1 (solid line): age, sex, PC1wbc, 9 CpGs methylation levels (cg01879420, cg06417478, cg20389709, cg13862711, cg25755428, cg06205333, cg25317025, cg23870316, cg00667948) on the following subgroups of MESO_ALL: (A) 107 participants, including 71 asbestos-exposed controls and 36 participants developing MPM within five years from recruitment (AUC model 0=0.656, AUC model 1=0.885, De-Long's $p=6.44x10^{-5}$); (B) 111 participants, including 71 asbestos-exposed controls and 40 participants who developed MPM in 5-10 years from recruitment (AUC model 0=0.671, AUC model 1=0.795, De-Long's p=0.018); (C) 129 participants, including 71 asbestos-exposed controls and 58 participants who developed MPM more than 10 years from recruitment (AUC model 0=0.627, AUC model 1=0.747, De-Long's p=0.018). AUC, area under the curve.

Fig. 2. Trend test analysis of nine differentially methylated CpGs on MESO_ALL cases divided into groups by time from recruitment to diagnosis, controlling for age, gender and PC1wbc: >10 years (n=58), between 5 and 10 years (n=40), <5 years (n=36). The table shows trend test p values among groups of cases. Boxplots of significant CpGs are shown on the right, including also DNAm values of the control group (n=71).

Table I. Sample characteristics within the two EPIC-Meso subgroups analysed: MESO_ALL and MESO_5YRS groups The description of MESO_ALL cases and MESO_5YRS cases divided by years from recruitment to MPM diagnosis is included. Percentages were calculated in relation to the whole subgroup (*) or the N individuals of each column (**).

	MESO_ALL		***MESO_ALL Cases***			***MESO_5YRS***		***MESO_5YRS Cases***		
	Cases	***Controls***	***>10 years***	***5-10 years***	***<5 years***	***Cases***	***Controls***	***3-5 years***	***1.5-3 years***	***< 1.5 years***
N (%*)	134 (50)	134 (50)	58 (43.3)	40 (29.9)	36 (26.7)	36 (50)	36 (50)	14 (38.9)	13 (36.1)	9 (25)
SEX (%**)										
M	107 (79.9)	107 (79.9)	47 (81)	30 (75)	30 (83.3)	30 (83.3)	30 (83.3)	12 (85.7)	10 (76.9)	8 (88.9)
F	27 (20.1)	27 (20.1)	11 (19)	10 (25)	6 (16.7)	6 (16.7)	6 (16.7)	2 (14.3)	3 (23.1)	1 (11.1)
AGE (mean, sd)	57.2, 6.9	57.2, 6.9	56.4, 5.8	57.6, 8.6	58.2, 6.6	58.2, 6.6	58.2, 6.7	59.1 (7.3)	55.7 (6.5)	60.5 (4.9)
ASBESTOS EXPOSURE (%**)										
0 (no)	21 (15.7)	21 (15.7)	7 (12.1)	8 (20)	6 (16.6)	6 (16.6)	6 (16.6)	2 (14.3)	2 (15.4)	2 (22.2)
1 (medium)	28 (20.9)	28 (20.9)	12 (20.7)	5 (12.5)	11 (30.6)	11 (30.6)	11 (30.6)	3 (21.4)	6 (46.2)	2 (22.2)
2 (high)	43 (32.1)	43 (32.1)	21 (36.2)	13 (32.5)	9 (25)	9 (25)	9 (25)	3 (21.4)	4 (30.8)	2 (22.2)
NA	42 (31.3)	42 (31.3)	18 (31)	14 (35)	10 (27.8)	10 (27.8)	10 (27.8)	6 (42.9)	1 (7.7)	3 (33.3)

Table II. Description of the 22 differentially methylated CpGs identified by 10-fold cross-validation EWAS analysis: number of runs for which conditions of significance exist, their genomic localization, methylation levels (β-values) in cases and controls, effect size range in cross-validation runs and the p value range among runs are indicated. Signals selected after correlation analysis are highlighted in bold.

CpG ID	runs (n) with eff.size >10% (abs value) and p-value <0.05	CHR	MAPINFO	Gene name	Relation to gene	β-value MPM cases (n=36)	β-value MPM controls (n=36)	Effect size min (min as abs value)	Effect size max (max as abs value)	p-value min	p-value max
cg23870316	**10**	**8**	**2216155**			**0.85**	**0.69**	**0.146**	**0.180**	**0.0001**	**0.002**
cg20389709	**10**	**2**	**10184303**	**KLF11**	**5'UTR;TSS200;Body**	**0.12**	**0.28**	**-0.123**	**-0.197**	**0.013**	**0.00005**
cg13498216	8	2	10184305	KLF11	5'UTR;TSS200;Body	0.08	0.21	-0.106	-0.144	0.00004	0.002
cg06417478	**9**	**19**	**12876846**	**HOOK2**	**Body**	**0.26**	**0.44**	**-0.159**	**-0.253**	**0.002**	**0.048**
cg23899408	10	19	12877188	HOOK2	Body	0.27	0.43	-0.141	-0.226	0.043	0.001
cg16474696	9	19	13875014	MRI1	TSS1500	0.33	0.2	0.108	0.166	0.0006	0.022
cg25755428	**10**	**19**	**13875111**	**MRI1**	**TSS1500**	**0.41**	**0.24**	**0.134**	**0.234**	**0.0001**	**0.023**
cg25317025	**9**	**18**	**47019823**	**RPL17**	**TSS1500**	**0.55**	**0.44**	**0.103**	**0.138**	**0.003**	**0.033**
cg00667948	**9**	**13**	**100651721**			**0.88**	**0.76**	**0.106**	**0.135**	**0.007**	**0.045**
cg01879420	**9**	**6**	**111194645**	**AMD1**	**TSS1500;5'UTR**	**0.17**	**0.29**	**-0.107**	**-0.145**	**0.00003**	**0.004**
cg18434912	8	6	111194786	AMD1	TSS1500;5'UTR	0.2	0.3	-0.103	-0.139	0.0002	0.010
cg06205333	**8**	**1**	**112161618**	**RAP1A**	**TSS1500**	**0.63**	**0.76**	**-0.109**	**-0.174**	**0.002**	**0.023**
cg04282082	9	9	124988720	LHX6	Body	0.67	0.51	0.140	0.244	0.0003	0.045
cg11479503	8	9	124989052	LHX6	Body	0.84	0.72	0.117	0.161	0.00002	0.002
cg21469772	9	9	124989294	LHX6	Body	0.74	0.59	0.124	0.224	0.00002	0.025
cg13571460	9	9	124989337	LHX6	Body	0.72	0.58	0.109	0.199	0.00002	0.024
cg05136264	9	9	124989408	LHX6	Body	0.73	0.6	0.109	0.193	0.0001	0.036
cg05037505	8	9	124989550	LHX6	Body	0.5	0.4	0.106	0.145	0.0006	0.019
cg15124400	9	9	124989839	LHX6	1stExon;5'UTR;Body	0.78	0.6	0.162	0.259	0.0002	0.021
cg13862711	**9**	**9**	**124989915**	**LHX6**	**Body**	**0.74**	**0.54**	**0.174**	**0.279**	**0.0001**	**0.017**
cg04622888	9	9	124990010	LHX6	TSS200;Body	0.7	0.55	0.141	0.217	0.0006	0.026
cg03363289	9	9	124990165	LHX6	Body	0.45	0.32	0.121	0.193	0.0002	0.021

Table III. ROC analysis for each of the nine statistically differentially methylated CpGs, performed on all MESO_5YRS cases (n=36) and all the asbestos-exposed controls (n=71). Model 0: age, sex, PC1wbc; model 1: age, sex, PC1wbc, nine CpGs methylation levels.

CpG ID	Gene name	AUC MODEL 0	AUC MODEL 1	De Long's p
cg01879420	AMD1	0.66	0.73	0.1
cg06417478	HOOK2	0.66	0.69	0.42
cg20389709	KLF11	0.66	0.74	0.08
cg13862711	LHX6	0.66	0.71	0.2
cg25755428	MRI1	0.66	0.71	0.2
cg06205333	RAP1A	0.66	0.67	0.51
cg25317025	RPL17	0.66	0.67	0.52
cg23870316		0.66	0.75	**0.02**
cg00667948		0.66	0.71	0.12

SUPPLEMENTARY MATERIALS

Blood cell DNA methylation biomarkers in preclinical malignant pleural mesothelioma: the EPIC prospective cohort.

A. Allione, C. Viberti, I. Cotellessa, C. Catalano, E. Casalone, G. Cugliari, A. Russo, S. Guarrera, D. Mirabelli, C. Sacerdote, M. Gentile, F. Eichelmann, M. B. Schulze, S. Harlid[11], A. K. Eriksen, A. Tjønneland, M. Andersson, M. E. T. Dollé[15], H. Van Puyvelde, E. Weiderpass, M. Rodriguez-Barranco, A. Agudo, A. K Heath, M. Chirlaque, T. Truong, D. Dragic, G. Severi, S. Sieri[25], T. M. Sandanger, E. Ardanaz, P. Vineis, G. Matullo.

Table of Contents:

Table S1. Asbestos exposure assessment: we built up a job-exposure matrix (JEM) assigning to each occupation an "exposure probability" and an "exposure intensity" based on occupational history data (occupational categories 1-52). Both "exposure probability" and an "exposure intensity" are coded as: 0 = no probability/intensity; 1 = low probability/intensity; 2 = intermediate probability/intensity; 3 = high probability/intensity.

Exposure Index / Cumulative Exposure Index: the two values (probability and intensity) are then used to build up, for each occupation, an "Exposure Index" given by "Probability of exposure" * "intensity of exposure". The Exposure Index may thus assume the following values: 0, 1, 2, 3, 4, 6, 9. Each occupation has its own Exposure Index, which may be assigned to each individual as his/her own Exposure Index. If a participant has several occupations, a "cumulative exposure index" is computed. The cumulative exposure index is the sum of all the Exposure Indexes (one for each occupation) of that individuals. For participants with just one occupation the Exposure Index and the Cumulative Exposure Index coincide.

Derivative Exposure Indexes: the Exposure Index assigned to each individual according to his/her occupation (or the Cumulative Exposure Index in the case of multiple occupations) is used to build up derivative exposure indexes: 1) a "Binary Exposure Index", coded as: 0 = no exposure (if the Exposure Index, or cum. index in the case of multiple jobs, is < 3); 1 = exposed (if the Exposure Index, or cum. index in the case of multiple jobs, is ≥ 3); 2) a "3 levels categorical var.", coded as: 0 = no exposure (if the Exposure Index, or cum. index in the case of multiple jobs, is = 0); 1 = low exposure (if the Exposure Index, or cum. index in the case of multiple jobs, is >0 and <4); 2 = high exposure (if the Exposure Index, or cum. index in the case of multiple jobs, is ≥ 4). The table shows the 52 occupational categories.

	EPIC list of occupations	Probability of exposure	Intensity of Exposure	Exposure index	Binary exposure index
1	1.1 Livestock breeding	0	0	0	0
2	1.2 Agriculture	0	0	0	0
3	2 mines or quarries	0	0	0	0
4	3 Foundry	1	2	2	0
5	- Steel	1	2	2	0
6	- Special alloys	1	2	2	0
7	4 Galvanic	0	0	0	0
8	5 Chemical Industry	1	2	2	0
9	- Refinery	1	2	2	0
10	- Dyes Production	1	2	2	0
11	- Chemical Laboratory	0	0	0	0
12	6 Rubber Industry	1	1	1	0
13	7 Textile Industry	1	1	1	0
14	- Of tissues Dyeing	2	2	4	1
15	- Weaving	1	2	2	0
16	8 Processing and tanning	0	0	0	0
17	9 Production of shoes and leather	0	0	0	0
18	10 Woodworking	0	0	0	0
19	- Production of furniture	0	0	0	0
20	11 Metalworking	1	1	1	0
21	- Turning, drilling, milling etc..	0	0	0	0
22	- Welding	2	2	4	1
23	- Painting	0	0	0	0
24	12 Boatyard	2	3	6	1
25	13 Electrical and Electronics Industry	1	1	1	0
26	14 Glass Industry	1	2	2	0
27	15 Typography	0	0	0	0
28	16 Construction	1	2	2	0
29	- Roof Waterproofing	2	2	4	1
30	- Asphalt	0	0	0	0
31	- Demolition	2	2	4	1
32	17 Transport	1	1	1	0
33	- Truck driver	1	2	2	0
34	- Driver	0	0	0	0
35	- Taxi driver	0	0	0	0
36	18 Nuclear Industry	1	2	2	0
37	19 Production of paper or cellulose	1	2	2	0
38	20 Production of asbestos and asbestos-cement	3	3	9	1
39	21 Worker with asbestos insulation	3	3	9	1
40	22 Production of cement	0	0	0	0
41	23 Production of ceramics	0	0	0	0
42	24 Butcher	0	0	0	0
43	25 Painter, painter	1	1	1	0
44	26 Welder	2	2	4	1
45	27 Hairdresser	0	0	0	0
46	28 Gas Station	0	0	0	0
47	29 Auto Mechanic	1	1	1	0
48	30 Bartender	0	0	0	0
49	31 Warden restaurant	0	0	0	0
50	32 Medical and Health Services	0	0	0	0
51	33 Electrician	2	2	4	1
52	34 Other	0	0	0	0

Table S2. Estimated WBCs percentages on MESO_ALL participants

WBCs (mean, sd)	Cases (n=134)	Controls (n=134)	Wilcoxon p-value
Basophils	1.55, 0.82	1.71, 0.86	0.27
B memory	2.75, 0.9	2.82, 0.78	0.42
B naïve	2.71, 1.4	2.92, 1.42	0.19
CD4 memory	8.22, 3.26	9.27, 3.48	**0.02**
CD4 naïve	3.88, 2.57	3.95, 2.88	0.85
CD8 memory	7.46, 5.33	8.2, 4.71	0.07
CD8 naïve	1.39, 1.18	1.44, 1.36	0.84
Eosinophils	1.43, 1.75	1.72, 2.04	0.17
Monocytes	7.49, 2.2	7.76, 2.12	0.42
Neutrophils	55.49, 10.57	52.53, 10.22	**0.04**
NK	4.74, 2.21	5.15, 2.25	0.12
T reg	1.1, 0.86	0.94, 0.75	0.18

Table S3. Estimated WBCs percentages on MESO_5YRS participants

WBCs (mean, sd)	Cases (n=36)	Controls (n=36)	Wilcoxon p-value
Basophils	1.68, 0.83	1.65, 0.74	0.8
B memory	2.56, 0.73	2.66, 0.82	0.78
B naïve	3.02, 1.58	2.81, 1.28	0.76
CD4 memory	8.07, 2.88	8.82, 3.32	0.46
CD4 naïve	3.44, 2.31	4.28, 2.49	0.16
CD8 memory	7.6, 5.25	6.87, 4.17	0.71
CD8 naïve	1.22, 0.98	1.47, 1.17	0.42
Eosinophils	1.69, 2.1	1.7, 1.91	0.75
Monocytes	7.31, 1.95	6.94, 1.58	0.42
Neutrophils	56.52, 10.2	55.19, 8.08	0.52
NK	4.21, 1.93	5.06, 2.33	0.08
T reg	1.09, 0.78	1.01, 0.76	0.75

Table S4. Pearson correlation analysis of the differentially methylated CpGs in MESO_5YRS groups. The statistically significant values of Pearson correlation test (rho > |0.7|, p ≤0.05) are highlighted in bold.

p value	gene	RAP1A	KLF11	KLF11	AMD1	AMD1		LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6		RPL17	HOOK2	HOOK2	MRI1	MRI1
gene	cg ID	cg06205333	cg20389709	cg13498216	cg01879420	cg18434912	cg23870316	cg04282082	cg11479503	cg21469772	cg13571460	cg05136264	cg05037505	cg15124400	cg13862711	cg04622888	cg03363289	cg00667948	cg25317025	cg06417478	cg23899408	cg16474696	cg25755428
RAP1A	cg06205333	0																					
KLF11	cg20389709	0.3324	0																				
KLF11	cg13498216	0.2221	**3.07E-21**	0																			
AMD1	cg01879420	0.0783	0.1627	0.0899	0																		
AMD1	cg18434912	0.0771	0.1679	0.0927	**6.44E-29**	0																	
	cg23870316	0.0435	0.0002	0.0001	0.0631	0.0609	0																
LHX6	cg04282082	0.0001	0.0565	0.0277	0.0013	0.0012	0.0710	0															
LHX6	cg11479503	0.0002	0.0344	0.0157	0.0015	0.0014	0.0525	**5.91E-25**	0														
LHX6	cg21469772	0.0001	0.0409	0.0190	0.0012	0.0012	0.0557	**1.52E-28**	**4.60E-29**	0													
LHX6	cg13571460	0.0001	0.0356	0.0167	0.0018	0.0017	0.0570	**6.87E-26**	**1.09E-33**	**9.77E-30**	0												
LHX6	cg05136264	0.0001	0.0510	0.0247	0.0014	0.0013	0.0621	**6.58E-32**	**1.68E-25**	**4.90E-31**	**6.67E-27**	0											
LHX6	cg05037505	0.0001	0.0581	0.0283	0.0012	0.0011	0.0618	**3.83E-25**	**8.54E-21**	**1.75E-24**	**2.05E-21**	**7.11E-27**	0										
LHX6	cg15124400	0.0001	0.0659	0.0322	0.0009	0.0008	0.0662	**3.00E-26**	**6.80E-22**	**7.02E-25**	**4.92E-22**	**1.61E-27**	**4.03E-27**	0									
LHX6	cg13862711	0.0001	0.0554	0.0266	0.0011	0.0011	0.0589	**6.13E-30**	**5.05E-24**	**2.06E-28**	**1.81E-24**	**3.32E-33**	**6.53E-28**	**5.99E-32**	0								
LHX6	cg04622888	0.0001	0.0771	0.0385	0.0011	0.0010	0.0668	**4.94E-24**	**4.69E-20**	**3.17E-22**	**2.87E-20**	**1.14E-24**	**6.19E-26**	**1.17E-30**	**1.25E-27**	0							
LHX6	cg03363289	0.0001	0.0562	0.0276	0.0015	0.0014	0.0687	**2.28E-27**	**1.61E-21**	**4.20E-25**	**2.29E-22**	**2.00E-27**	**3.55E-29**	**1.00E-24**	**6.41E-27**	**4.48E-24**	0						
	cg00667948	0.3601	0.9111	0.8353	0.0085	0.0084	0.5930	0.4992	0.5140	0.4566	0.5156	0.4879	0.4317	0.4478	0.4792	0.4868	0.4719	0					
RPL17	cg25317025	0.2281	0.0232	0.0208	0.2557	0.2820	0.1394	0.3962	0.3528	0.3504	0.3516	0.3784	0.3884	0.4131	0.3968	0.4730	0.3921	0.3634	0				
HOOK2	cg06417478	0.1831	0.0889	0.0780	0.1648	0.1476	0.0229	0.0105	0.0089	0.0099	0.0077	0.0091	0.0121	0.0113	0.0098	0.0110	0.0122	0.7333	0.5001	0			
HOOK2	cg23899408	0.1603	0.0832	0.0725	0.1543	0.1378	0.0179	0.0109	0.0093	0.0102	0.0081	0.0094	0.0124	0.0118	0.0102	0.0114	0.0126	0.7804	0.4658	**1.19E-30**	0		
MRI1	cg16474696	0.8429	0.0158	0.0170	0.0057	0.0057	0.1096	0.3117	0.2672	0.2959	0.2760	0.3255	0.3760	0.3574	0.3354	0.3944	0.3580	0.3125	0.3509	0.0316	0.0303	0	
MRI1	cg25755428	0.4679	0.0297	0.0354	0.0248	0.0246	0.1711	0.6471	0.5729	0.6242	0.5870	0.6668	0.7482	0.7151	0.6824	0.7657	0.7208	0.4629	0.4135	0.0491	0.0479	**8.04E-18**	0

RHO	gene	RAP1A	KLF11	KLF11	AMD1	AMD1		LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6	LHX6		RPL17	HOOK2	HOOK2	MRI1	MRI1
gene	cg ID	cg06205333	cg20389709	cg13498216	cg01879420	cg18434912	cg23870316	cg04282082	cg11479503	cg21469772	cg13571460	cg05136264	cg05037505	cg15124400	cg13862711	cg04622888	cg03363289	cg00667948	cg25317025	cg06417478	cg23899408	cg16474696	cg25755428
RAP1A	cg06205333	1.000																					
KLF11	cg20389709	0.057	1.000																				
KLF11	cg13498216	0.070	**0.944**	1.000																			
AMD1	cg01879420	0.109	0.107	0.147	1.000																		
AMD1	cg18434912	0.117	0.106	0.154	**0.961**	1.000																	
	cg23870316	-0.176	-0.303	-0.372	-0.120	-0.118	1.000																
LHX6	cg04282082	-0.244	-0.040	-0.093	-0.211	-0.213	0.107	1.000															
LHX6	cg11479503	-0.196	-0.073	-0.129	-0.164	-0.160	0.157	**0.946**	1.000														
LHX6	cg21469772	-0.241	-0.072	-0.123	-0.201	-0.191	0.154	**0.976**	**0.978**	1.000													
LHX6	cg13571460	-0.209	-0.084	-0.132	-0.161	-0.152	0.134	**0.964**	**0.983**	**0.984**	1.000												
LHX6	cg05136264	-0.236	-0.056	-0.087	-0.203	-0.191	0.143	**0.977**	**0.953**	**0.982**	**0.971**	1.000											
LHX6	cg05037505	-0.289	-0.069	-0.099	-0.243	-0.232	0.145	**0.950**	**0.892**	**0.940**	**0.919**	**0.962**	1.000										
LHX6	cg15124400	-0.207	-0.020	-0.065	-0.254	-0.240	0.155	**0.963**	**0.941**	**0.964**	**0.948**	**0.974**	**0.959**	1.000									
LHX6	cg13862711	-0.238	-0.044	-0.092	-0.234	-0.221	0.162	**0.980**	**0.950**	**0.979**	**0.962**	**0.982**	**0.960**	**0.988**	1.000								
LHX6	cg04622888	-0.211	-0.008	-0.054	-0.259	-0.241	0.164	**0.952**	**0.913**	**0.947**	**0.925**	**0.958**	**0.959**	**0.979**	**0.978**	1.000							
LHX6	cg03363289	-0.263	-0.055	-0.091	-0.209	-0.186	0.121	**0.954**	**0.890**	**0.939**	**0.911**	**0.948**	**0.957**	**0.925**	**0.947**	**0.940**	1.000						
	cg00667948	-0.057	0.077	0.064	-0.241	-0.239	-0.112	0.093	0.117	0.152	0.119	0.125	0.148	0.139	0.121	0.108	0.123	1.000					
RPL17	cg25317025	-0.053	-0.123	-0.144	-0.016	0.016	0.131	0.143	0.171	0.172	0.172	0.161	0.131	0.138	0.141	0.087	0.141	0.162	1.000				
HOOK2	cg06417478	0.085	0.184	0.174	0.095	0.117	-0.140	-0.093	-0.072	-0.083	-0.101	-0.103	-0.101	-0.092	-0.106	-0.108	-0.076	0.156	0.066	1.000			
HOOK2	cg23899408	0.115	0.193	0.173	0.101	0.128	-0.165	-0.086	-0.067	-0.079	-0.094	-0.097	-0.098	-0.083	-0.097	-0.101	-0.079	0.137	0.057	**0.991**	1.000		
MRI1	cg16474696	0.156	-0.157	-0.158	-0.245	-0.229	0.137	0.194	0.217	0.198	0.213	0.176	0.130	0.146	0.160	0.117	0.163	0.164	0.130	-0.120	-0.121	1.000	
MRI1	cg25755428	0.219	-0.148	-0.142	-0.207	-0.196	0.123	0.122	0.175	0.136	0.162	0.117	0.048	0.079	0.100	0.049	0.080	0.111	0.129	-0.136	-0.133	0.935	1.000

Table S5. Differentially methylated regions (DMRs), with mean of the estimate $>|0.5|$ ("meandiff") between cases and controls in MESO_5YRS participants. DMRs were calculated by ChAMP approach (champ DMR function).

DMR	Chr	start	end	width	strand	no.cpgs	min_smoothed_fdr	Stouffer	HMFDR	Fisher	maxdiff	meandiff	overlapping.genes
DMR_971	chr9	124988720	124990785	2066	*	16	3.58E-13	0.933	0.523	0.958	-0.196	-0.101	LHX6
DMR_469	chr8	637468	638330	863	*	5	5.02E-03	0.637	0.558	0.832	-0.143	-0.080	ERICH1
DMR_809	chr3	195488922	195490309	1388	*	9	5.83E-04	0.696	0.566	0.924	-0.129	-0.076	MUC4
DMR_210	chr17	20799470	20799532	63	*	4	6.83E-03	0.565	0.530	0.751	-0.088	-0.069	RP11-344E13.3
DMR_511	chr8	58055876	58056175	300	*	4	3.74E-02	0.691	0.594	0.844	-0.079	-0.058	RP11-513O17.2
DMR_71	chr22	49843259	49843754	496	*	3	2.17E-02	0.544	0.524	0.694	-0.084	-0.056	C22orf34
DMR_631	chr19	37825211	37826008	798	*	10	2.93E-07	0.592	0.523	0.883	0.062	0.042	HKR1
DMR_1054	chr12	130822256	130822605	350	*	7	5.35E-02	0.866	0.657	0.970	0.063	0.044	PIWIL1
DMR_391	chr16	86795398	86796158	761	*	4	2.39E-04	0.810	0.545	0.808	0.089	0.045	NA
DMR_626	chr17	170770	172159	1390	*	12	3.76E-09	0.541	0.508	0.882	0.071	0.045	RP11-1260E13.1, RPH3AL
DMR_748	chr21	45160229	45161603	1375	*	7	3.88E-03	0.731	0.574	0.908	0.078	0.046	PDXK
DMR_1008	chr2	206628088	206629314	1227	*	12	1.76E-06	0.881	0.568	0.964	0.086	0.048	NRP2, AC007362.3
DMR_301	chr1	153581946	153582683	738	*	4	7.32E-03	0.643	0.542	0.785	0.105	0.061	S100A16
DMR_124	chr1	112161618	112161864	247	*	3	9.38E-03	0.589	0.538	0.723	0.133	0.063	RAP1A
DMR_87	chr19	12876846	12877188	343	*	4	2.37E-03	0.508	0.504	0.705	0.183	0.176	HOOK2

Table S6. Trend test, adjusted for age, sex and PC1wbc, of 22 CpGs DNA methylation in 3 asbestos-exposure groups in MESO_5YRS individuals (26 cases and 26 controls with exposure assessment).

CASES									
ID	**linear_estimate**	**std_err**	**pval**	**exp0_mean**	**exp0_sd**	**exp1_mean**	**exp1_sd**	**exp2_mean**	**exp2_sd**
cg04282082	0.0322	0.11	0.76	0.68	0.34	0.63	0.19	0.74	0.26
cg04622888	0.002	0.1	0.98	0.69	0.36	0.68	0.22	0.78	0.21
cg18434912	0.0231	0.05	0.67	0.17	0.07	0.26	0.17	0.19	0.07
cg01879420	0.0363	0.05	0.47	0.16	0.06	0.22	0.16	0.17	0.06
cg03363289	0.0319	0.08	0.7	0.47	0.27	0.4	0.15	0.52	0.22
cg21469772	0.0039	0.08	0.96	0.75	0.28	0.74	0.16	0.79	0.21
cg23899408	-0.0222	0.1	0.82	0.23	0.24	0.28	0.24	0.27	0.21
cg23870316	-0.012	0.06	0.83	0.79	0.26	0.86	0.1	0.84	0.13
cg16474696	0.0061	0.08	0.94	0.44	0.17	0.29	0.23	0.32	0.22
cg05037505	0.0159	0.07	0.83	0.51	0.23	0.48	0.14	0.54	0.18
cg06205333	-0.1484	0.08	0.09	0.72	0.21	0.68	0.19	0.55	0.23
cg15124400	0.0126	0.1	0.9	0.75	0.37	0.78	0.21	0.84	0.22
cg25317025	0.0225	0.09	0.8	0.48	0.2	0.53	0.23	0.53	0.17
cg05136264	0.0262	0.08	0.75	0.73	0.27	0.71	0.16	0.79	0.18
cg13862711	0.0191	0.11	0.86	0.72	0.38	0.72	0.23	0.81	0.23
cg00667948	0.0756	0.08	0.34	0.81	0.29	0.91	0.02	0.86	0.2
cg06417478	-0.0293	0.11	0.8	0.23	0.3	0.26	0.25	0.26	0.26
cg13498216	0.0394	0.02	0.13	0.05	0.02	0.07	0.07	0.09	0.07
cg20389709	0.0514	0.04	0.24	0.06	0.02	0.14	0.14	0.12	0.09
cg11479503	0.002	0.05	0.97	0.84	0.18	0.84	0.11	0.88	0.12
cg25755428	-0.0246	0.1	0.8	0.6	0.21	0.36	0.33	0.4	0.22
cg13571460	0.0166	0.07	0.82	0.73	0.24	0.72	0.13	0.77	0.18
CONTROLS									
ID	**linear_estimate**	**std_err**	**pval**	**exp0_mean**	**exp0_sd**	**exp1_mean**	**exp1_sd**	**exp2_mean**	**exp2_sd**
cg04282082	-0.0039	0.11	0.97	0.55	0.29	0.47	0.25	0.51	0.25
cg04622888	0.05	0.11	0.65	0.56	0.28	0.5	0.28	0.58	0.21
cg18434912	-0.0294	0.07	0.66	0.27	0.13	0.33	0.19	0.26	0.13
cg01879420	-0.0071	0.06	0.91	0.22	0.08	0.32	0.19	0.25	0.13
cg03363289	0.0364	0.08	0.66	0.34	0.21	0.27	0.2	0.35	0.2
cg21469772	-0.0027	0.09	0.98	0.62	0.23	0.57	0.2	0.58	0.21
cg23899408	-0.1104	0.12	0.37	0.58	0.29	0.39	0.27	0.42	0.29
cg23870316	0.011	0.08	0.89	0.71	0.15	0.7	0.25	0.69	0.2
cg16474696	-0.045	0.06	0.42	0.22	0.1	0.2	0.15	0.18	0.14
cg05037505	0.0383	0.07	0.59	0.41	0.18	0.35	0.18	0.44	0.14
cg06205333	0.1272	0.08	0.14	0.64	0.29	0.75	0.17	0.84	0.14
cg15124400	0.0355	0.12	0.77	0.63	0.32	0.54	0.3	0.63	0.22
cg25317025	0.0379	0.07	0.57	0.41	0.16	0.43	0.22	0.53	0.17
cg05136264	0.0187	0.08	0.82	0.6	0.22	0.56	0.2	0.61	0.18
cg13862711	0.0164	0.12	0.89	0.58	0.31	0.49	0.29	0.56	0.26
cg00667948	-0.1071	0.09	0.27	0.88	0.08	0.73	0.23	0.72	0.27
cg06417478	-0.115	0.14	0.41	0.59	0.33	0.38	0.31	0.43	0.33
cg13498216	0.0243	0.06	0.7	0.17	0.17	0.2	0.14	0.21	0.14
cg20389709	0.028	0.09	0.75	0.26	0.28	0.27	0.19	0.28	0.19
cg11479503	6.00E-04	0.06	0.99	0.74	0.16	0.72	0.16	0.72	0.14
cg25755428	-0.01	0.08	0.9	0.22	0.12	0.26	0.2	0.22	0.2
cg13571460	0.0038	0.08	0.96	0.61	0.21	0.56	0.18	0.59	0.16

Table S7. ROC analysis of cg23870316 in MESO_ALL participants diagnosed with MPM at different times from recruitment, compared with asbestos-exposed controls. Model 0: age, sex, PC1wbc; and model 1: age, sex, PC1wbc, cg23870316 methylation

Time from diagnosis	Cases (n)	Controls (n)	AUC Model 0	AUC Model 1	De Long's p
<5 years	36	107	0.66	0.75	0.02
5-10 years	40	107	0.67	0.67	0.95
>10 years	58	107	0.63	0.67	0.38

Table S8. Trend test analysis of nine differentially methylated CpGs in cases who developed MPM within five years, controlling for age, gender and WBC's first principal component. MPM cases were divided into three groups by time from recruitment to diagnosis: <1.5 years, from 1.5 to 3 years, > 3 years. The table shows trend test p values for each CpG.

cg ID	gene name	linear effect size	pval
cg01879420	AMD1	0.041	0.192
cg23870316		-0.037	0.292
cg06205333	RAP1A	0.067	0.331
cg25317025	RPL17	0.046	0.419
cg13862711	LHX6	-0.072	0.367
cg00667948		-0.039	0.423
cg06417478	HOOK2	-0.030	0.690
cg20389709	KLF11	-0.019	0.616
cg25755428	MRI1	-0.069	0.350

Table S9. Differential DNAm in MESO_5YRS participants of diagnostic CpGs previously identified in MPM patients in retrospective studies, controlling for age, gender and PC1wbc. nd= not detected.

Biomarkers from Guarrera et al. J. Thorac. Oncol. 2019				
CpG ID	**Gene name**	**Relation to gene**	**Effect size**	**p-value**
cg04572930	FOXK1	Body	nd	nd
cg04739200	MYB	Body	nd	nd
cg01521397	TAF4	Body	-0.030	**0.011**
cg08450017	FYCO1	Body	0.013	0.104
cg26033526	TAP1	Body	nd	nd
cg23825480	MORC2	Body	0.006	0.449
cg00446123	LIME1	TSS200	0.007	0.493
Biomarkers from Cugliari et al. Cancers 2021				
cg03546163	FKBP5	5' UTR	-0.002	0.881
cg06633438	MLLT1	Body	-0.011	0.410

Figure S1. Flow diagram depicting the different EPIC-Meso subgroups analysed in the present study and relative analyses performed.

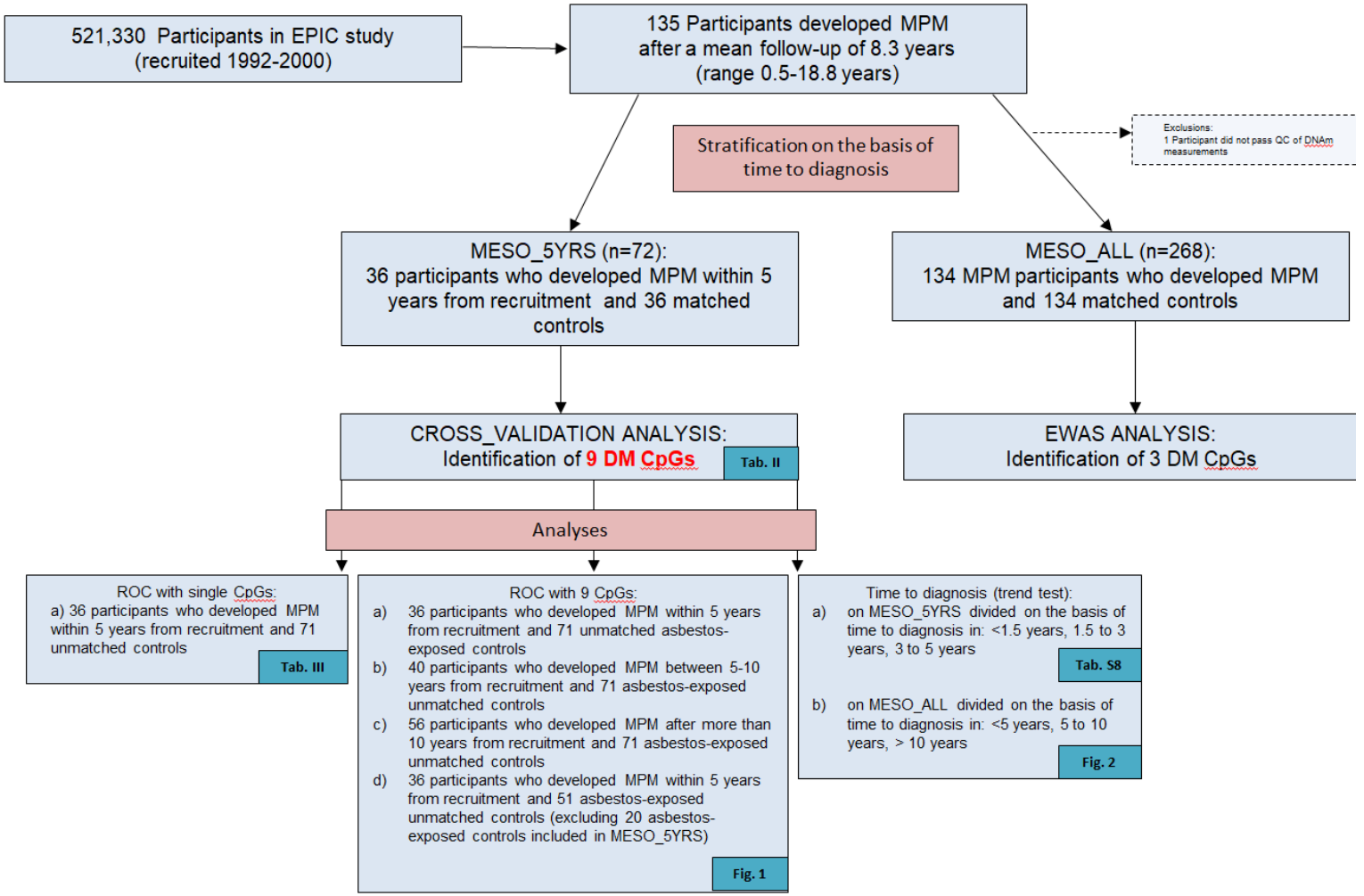


Fig. S2 Unsupervised hierarchical clustering heatmap including the 22 differentially methylated CpGs (see Table II) in MESO_5YRS group.

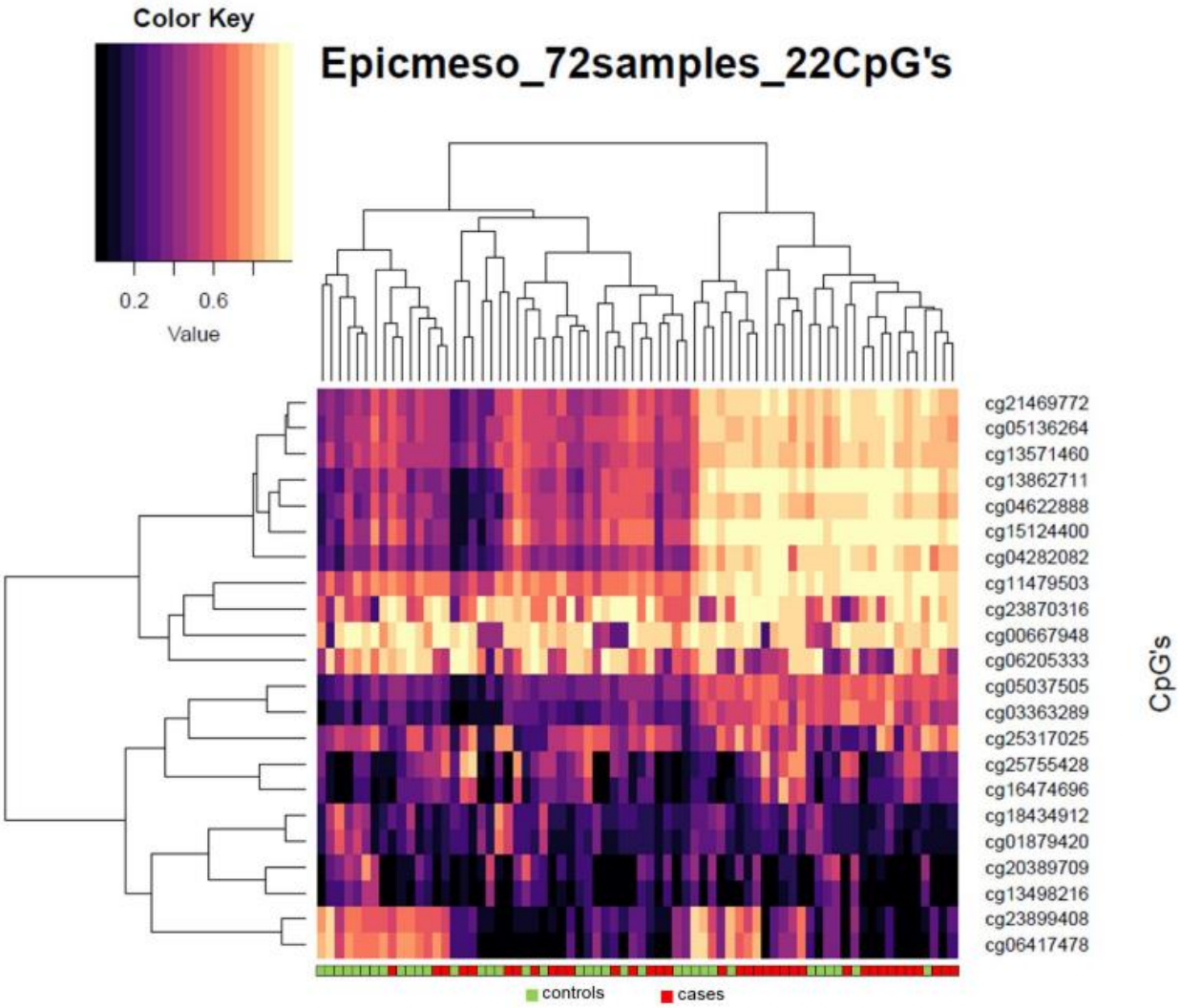
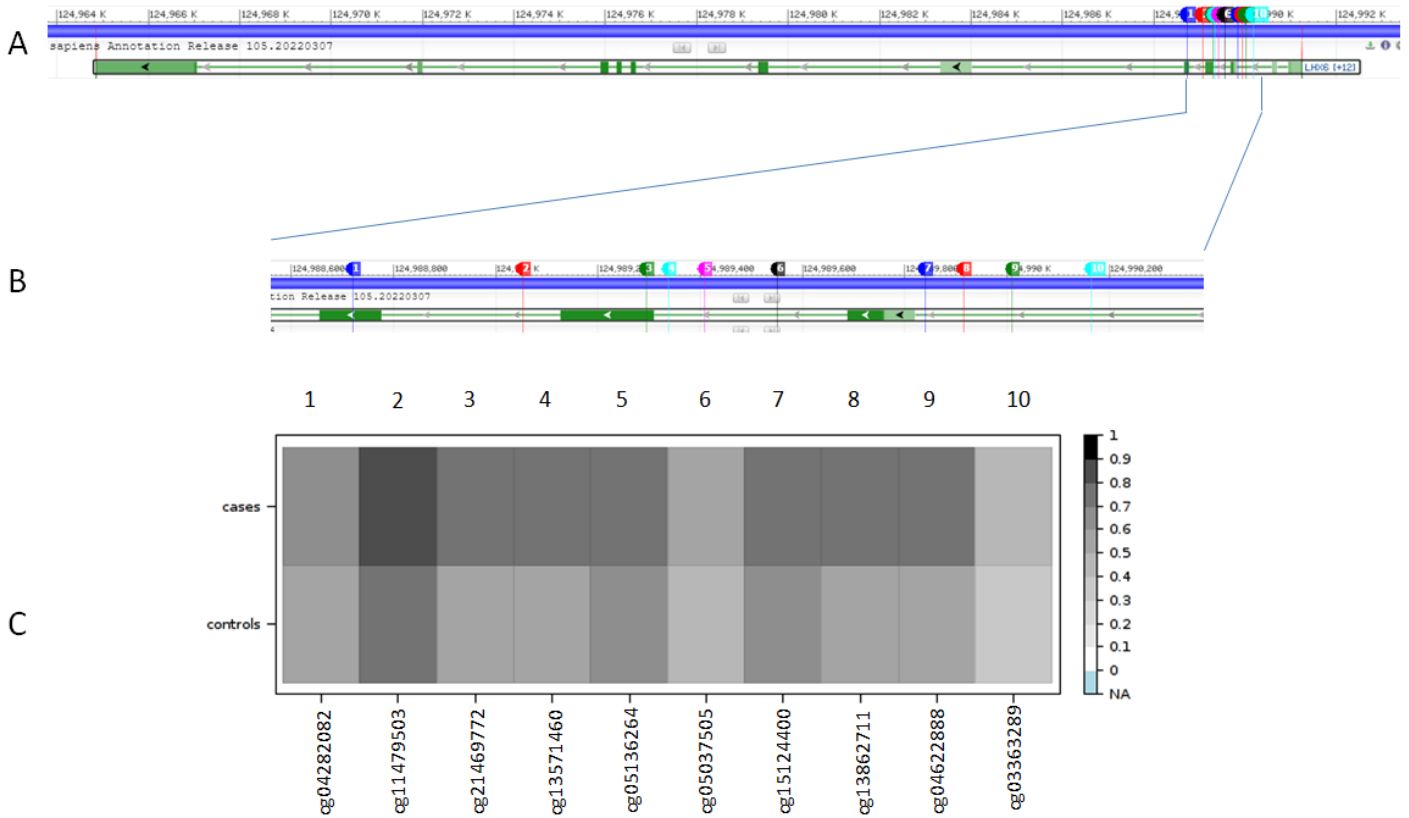
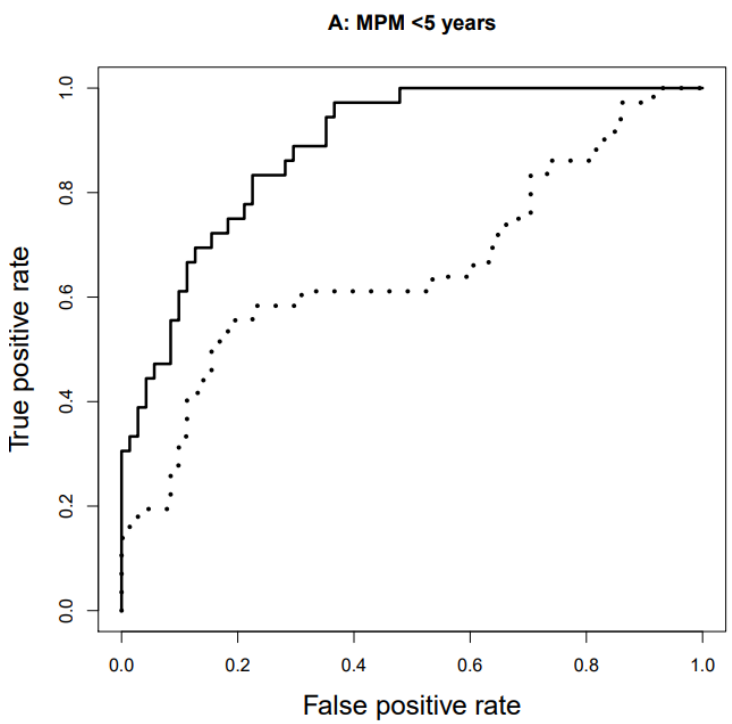
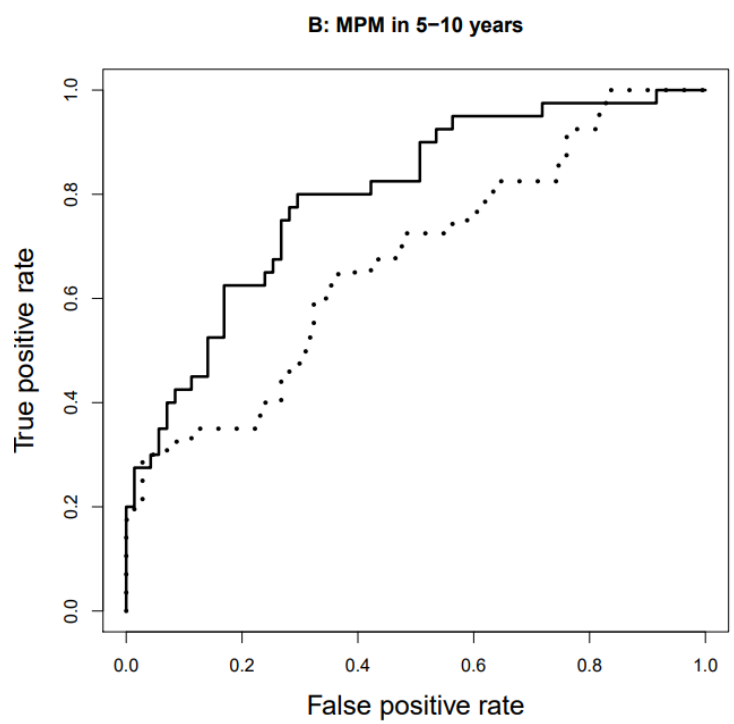
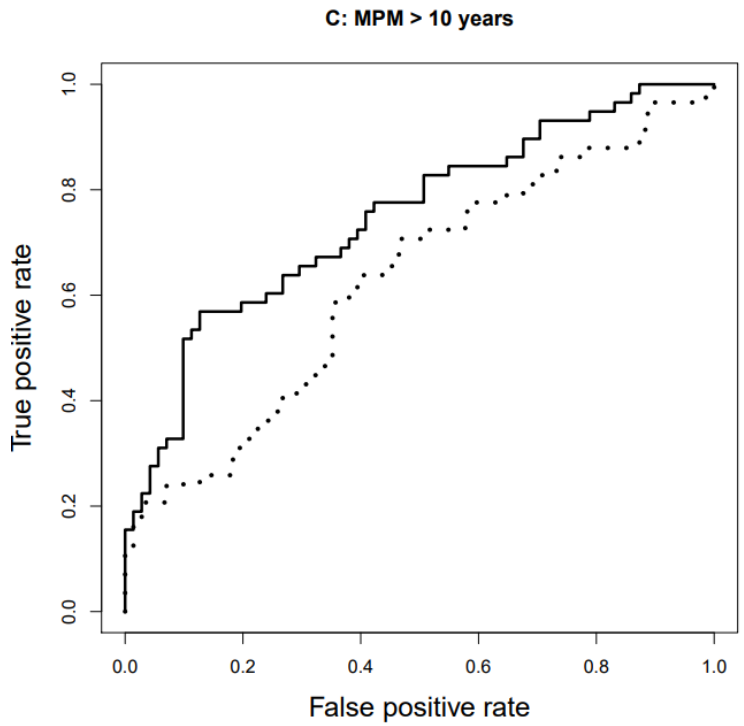


Figure S3. Gene-based visualization of ten differentially methylated CpG in LHX6 gene between cases and controls of MESO_5YRS individuals. (A) Gene location of CpGs (Genome Data Viewer, GDV, https://www.ncbi.nlm.nih.gov/genome/gdv/); (B) enlarged visualization of CpGs position; (C) Differential methylation visualization (Methylation Plotter, http://maplab.imppc.org/methylation_plotter/; Methylation plotter: a web tool for dynamic visualization of DNA methylation data Source Code for Biology and Medicine 2014, 9:11).




A: MPM <5 years
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


B: MPM in 5−10 years
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


C: MPM > 10 years
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

cg ID	gene name	linear effect size	pval
cg01879420	**AMD1**	**-0.060**	**0.006**
cg23870316		**0.107**	**0.001**
cg06205333	RAP1A	-0.059	0.055
cg25317025	RPL17	-0.001	0.968
cg13862711	LHX6	0.071	0.089
cg00667948		0.023	0.292
cg06417478	HOOK2	-0.021	0.571
cg20389709	KLF111	-0.029	0.181
cg25755428	**MRI1**	**0.088**	**0.021**