TITLE PAGE:

Title: Mesothelin promoter variants are associated with increased Soluble-Mesothelin Related Peptide (SMRP) levels in asbestos-exposed individuals

Authors: Chiara De Santi[1], Perla Pucci[2], Alessandra Bonotti[3], Ombretta Melaiu[4], Monica Cipollini[2], Elisa Barone[2], Elisa Paolicchi[2], Alda Corrado[2], Irene Lepori[2], Irene Dell'Anno[2], Lucia Pelle'[2], Luciano Mutti[5], Rudy Foddis[6], Alfonso Cristaudo[6], Federica Gemignani[2] and Stefano Landi[2,*].

Affiliation:

1. Respiratory Research Division, Department of Medicine, Education and Research Centre, Royal College of Surgeons in Ireland, Beaumont Hospital, Dublin 9, Ireland.
2. Department of Biology, University of Pisa, Via Derna 1, Pisa, Italy.
3. Preventive and Occupational Medicine, University Hospital of Pisa, Pisa, Italy.
4. Immuno-Oncology Laboratory, Department of Paediatric Haematology/Oncology, Ospedale Pediatrico Bambino Gesù, Viale di S. Paolo 15, 00146 Rome, Italy.
5. School of Environment and Life Sciences, University of Salford, Manchester, United Kingdom.
6. Department of Translational Research and of new Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy.

*Correspondence to: Stefano Landi, Department of Biology, University of Pisa, Via Derna, 1, 56126 Pisa, Italy. Tel.: +39 050 2211528; Fax: +39 0502211527, stefano.landi@unipi.it.

ABSTRACT

Soluble-Mesothelin Related Peptide (SMRP) is a promising diagnostic biomarker for malignant pleural mesothelioma (MPM), but various confounders hamper its usefulness in surveillance programs. We previously showed that a single nucleotide polymorphism (SNP) within the 3'untranslated region (3'UTR) of *mesothelin* (*MSLN*) gene could affect the levels of SMRP. Here, we focused on SNPs located within *MSLN* promoter and found a strong association between serum SMRP and variant alleles of rs3764247, rs3764246 (that is in strong linkage disequilibrium with rs2235504), and rs2235503 in non-MPM subjects. The inclusion of the genotype information led to an increase in SMRP specificity from 79.9% to 85.5%. Although not statistically significant, the MPM population showed the same trend of association. In order to study the biological role of these SNPs, the promoter region of *MSLN* was cloned upstream a reporter gene and the four most common haplotypes were compared in a dual luciferase assay. Rs3764247 was shown to have a functional role itself. The other SNPs were shown to interact with each other in a more complex way. Altogether, these data support the idea that SMRP performance is affected by individual (i.e. genetic) variables and that *MSLN* expression is influenced by SNPs located within the promoter regulatory region.

KEY WORDS: Malignant Pleural Mesothelioma, Single nucleotide polymorphism, soluble-mesothelin related peptide, promoter, biomarker specificity

INTRODUCTION

Mesothelin (MSLN) is a membrane-bound glycoprotein physiologically expressed by the mesothelial tissues of pleura, peritoneum, and pericardium (Hassan et al., 2004). Although its biological function is still unknown (Bera and Pastan, 2000), many types of cancer, including malignant pleural mesothelioma (MPM), show increased expression of MSLN compared to their non-malignant counterparts (Hassan and Ho, 2008). MPM is a highly aggressive tumor of the pleural cavities, associated with asbestos exposure and characterized by challenging diagnosis and poor prognosis (Panou et al., 2015). In recent years, several research groups suggested that MSLN could be helpful in the management of MPM, both as diagnostic tool (Robinson et al., 2003; Cristaudo et., 2007) and putative therapeutic target (Hassan et al., 2010; Hassan et al., 2014). In particular, high levels of the soluble form of MSLN, the so-called SMRP (Soluble Mesothelin-Related Peptides), were repeatedly observed in serum samples of MPM patients when compared to various types of control groups (Cristaudo et al., 2007; Pass et al., 2008; Fukuoka et a., 2013). Nonetheless, in spite of the initial findings, the real usefulness of SMRP within surveillance programs is hampered by a relatively high rate of false negatives as well as false positives (Cui et al., 2014). Various demographic and clinical variables were reported as possible confounders, such as body mass index, age, glomerular filtration rate, and lung function (Hollevoet et al., 2009; Park et al., 2010; Filiberti et al., 2013). Genetic factors were also shown to affect SMRP levels in non-MPM subjects. Thus, the inclusion of individuals' genetic information could improve the ROC curves calculation leading to slight improvements of the performance of SMRP as biomarker (Garritano et al., 2014). Previously, studying a broad cohort of non-MPM subjects, we reported an association between serum SMRP levels and rs1057147, a single nucleotide polymorphism (SNP) located within the 3'untranslated region (3'UTR) of *MSLN*. This SNP lies within the binding site for miR-611, thereby affecting the post-transcriptional regulation

of *MSLN* mRNA (Garritano et al., 2014). Similarly, genetic variants located within the promoter region of *MSLN* were found to be associated with SMRP levels in a small group of non-MPM volunteers (Cristaudo et al., 2011). Healthy subjects carrying the variant allele of rs3764247 A>C (reported as New1 in the original publication) showed increased SMRP levels compared to those carrying the AA genotype (Cristaudo et al., 2011). This could be ascribed to a different regulatory pattern depending on the presence of the variant or common allele. In the present work we analysed a large sample set and we were able to replicate the association between rs3764247 and SMRP levels. Moreover, in order to further explore the role of genetic variants in MSLN/SMRP regulation, we (i) evaluated the association between SMRP and other SNPs located within the proximal *MSLN* promoter and (ii) performed an *in vitro* study to assess the biological role of the selected SNPs. Altogether, these findings could help to refine the use of SMRP as diagnostic biomarker and to shed some light in the regulatory mechanisms of *MSLN* gene.

MATERIALS AND METHODS

SNPs selection

In the pilot study, an association between rs3764247 and SMRP was found (Cristaudo et al., 2011). Here, the association analysis was extended to other SNPs lying within the region of the proximal promoter of *MSLN*. Thus, selection criteria for the SNPs were: (i) to lie within 1000 bp (arbitrarily chosen) upstream the *MSLN* transcriptional start site (TSS); (ii) the frequency of the rare allele must be >0.05; (iii) to be reported as associated with *MSLN* mRNA expression in 278 lung tissue samples according to GTex portal (http://www.gtexportal.org/home/; Lonsdale et al., 2013). The linkage disequilibrium (LD) between the selected SNPs (i.e. rs3764247 A>C, rs3764246 A>G, rs2235503 C>A, rs2235504 A>G) and the most common haplotypes was estimated with HaploView software

version 4.2 (https://www.broadinstitute.org/.../haploview/haploview) using the TSI (Tuscans in Italy) population (however, CEU samples gave overlapping results).

Population description and genotyping

A total of 689 non-MPM subjects (healthy individuals n=371, or patients affected by benign respiratory diseases, BRDs, n=318) and 70 MPM volunteers were recruited at the University Hospital of Pisa as part of an occupational surveillance program on workers previously exposed to asbestos, as described in detail in Garritano et al., 2014. Table 1 shows the clinical and demographic characteristics of the sample set. The study was approved by the institutional ethical committee of the University Hospital of Pisa. All subjects gave written informed consent. For genotyping, whole blood and serum samples were obtained by venipuncture and kept at -80°C until examination. DNA was extracted from whole blood samples using EuroGOLD Blood DNA Mini Kit (EuroClone, Pero, Italy). Genotyping of the three selected SNPs (i.e. rs3764247, rs3764246 and rs2235503) was performed using KASPar® PCR SNP genotyping system (LGC Genomics Ltd, Teddington, Middlesex, UK) with a success rate >96%. Allele frequencies (shown in Table 1) were in agreement with those reported in HapMap project for TSI (0.20, 0.25, and 0.15 for rs3764247, rs3764246 and rs2235503, respectively) and followed the Hardy-Weinberg equilibrium (P=0.753, P=0.583, and P=0.625, respectively). Serum SMRP levels were measured using an enzyme-linked immuno-sorbent assay according to the manufacturer's instructions (Mesomark, Fujirebio Diagnostics, Japan).

Association analyses between genotypes and SMRP levels

In order to verify the association between genotypes and serum SMRP levels, one-way analysis of variance (ANOVA) was performed, stratified for health status (healthy, BRD, MPM), for each SNP. Tukey's multiple comparison tests were performed to assess pairwise differences between the three genotypes within each group. In order to ascertain the global

role of these SNPs in the association with SMRP in the different diagnostic groups, both the "non-MPM" (healthy subjects + BRD patients) and the MPM groups were stratified according to a three-SNPs classifier. According to this classifier, individuals carrying the common homozygote genotype for all the SNPs were considered as the reference category and were referred as carriers of the "L genotype" (L=low expression), whereas all the remaining subjects (i.e. carriers of at least one variant allele in one of the three SNPs) were considered to carry the "H genotype" (H=high). Then, a multivariate analysis of variance (mANOVA) was carried out to assess the association between SMRP values and L/H genotypes for each diagnostic group. The statistical significance threshold was set at 0.05 for all the analyses, which were performed using and StatGraphics Centurion XVI software (Manugistic, CA, USA).

Receiver Operating Characteristic (ROC) curves were generated with MedCalc statistical software (version 12.7.2.0, MedCalc Software, Belgium) comparing the non-MPM group versus the MPM group. First, the ROC curves were calculated without taking into account the genotypes. Then the curves were recalculated using SMRP levels of alternatively non-MPM volunteers carrying L or H genotype, versus the whole group of MPM patients (this group was not split in H/L genotype because no statistically significant differences were found in MPM patients).

Plasmids construction

The putative human *MSLN* promoter from nucleotides -1 to -1073 relative to the transcription start site was amplified by Q5® High-Fidelity DNA Polymerase (NEB, Ipswich, USA). As template, an individual carrying the common homozygote genotype for all the SNPs in study was selected from our sample set. The resultant polymerase chain reaction (PCR) amplicon was subsequently cloned into the XhoI site of the pGL3-basic vector (Promega, Madison, WI) using CloneEZ® PCR Cloning Kit (GenScript, Piscataway, USA). This construct,

bearing the most common haplotype in the TSI population, is from now on referred as "pGL3_HAP1". Subsequent site-directed mutagenesis reactions were performed to generate the other haplotype-mimicking plasmids with QuikChange II Site-Directed Mutagenesis Kits (Agilent, Santa Clara, CA, USA). The fidelity of the resulting constructs (pGL3_HAP1/2/3/4) was confirmed by sequencing, using the pGL3 external primers (pGL3_F and pGL3_R). The sequence of cloning, mutagenesis, and sequencing primers are reported in Supplementary Table 1.

Cell culture and luciferase reporter assays

Non-malignant transformed human pleural mesothelial cells (Met-5A) (Ke et al., 1989) were purchased from ATTC (American Type Tissue Collection) and cultured in Medium 199 (Gibco in Life Technologies) supplemented with 10% FBS, 1% pen/strep, 3 nM epidermal growth factor, 400 nM hydrocortisone, and 870 nM insulin. Met-5A cells were maintained in a humidified incubator at 37°C in 5% CO_2.

In three independent experiments, Met-5A cells were seeded in 24-wells plates at a final density of 50,000 cells/well and incubated for 24 hours. Cells were then transfected at 60-80% confluence with 400 ng of pGL3_HAP1/2/3/4 and 10 ng of internal control vector pRL-SV40 (Promega, Madison, USA) using Attractene reagent (Qiagen, Hilden, Germany). Twenty-four hours after transfection, a Dual-Luciferase Reporter Assay (Promega) was performed. Relative luciferase units (RLU) were expressed as mean value of the firefly luciferase/Renilla luciferase ratio of three independent experiments.

Functional annotation of the SNPs of interest

To assess the possible functional role of the SNPs of interest, we used the ENCODE-based tool HaploReg v4 (www.broadinstitute.org/mammals/haploreg) and RegulomeDB (http://www.regulomedb.org/). Overall, SNPs were analysed for mapping within DNase I

hypersensitive sites (DHSs), regulatory elements (enhancers and promoters), regulatory protein binding sites and altered motifs of transcription factors.

RESULTS

SNPs selection

In order to identify the genetic variants within *MSLN* proximal promoter (~1000 bp upstream to TSS) to be studied in association with SMRP, we searched for all SNPs significantly associated with *MSLN* mRNA expression in lung tissues on GTex portal (pleural tissues were unavailable). We found 86 cis-eQTLs with P-values ranging from $4.9x10^{-6}$ to $4.4x10^{-33}$. The region spanning *MSLN* TSS shows the highest associated SNPs. Table 2 lists the top ten associated SNPs with their main features. Among the 86 associated SNPs, we selected those lying within the 1000 bp upstream to TSS, i.e. rs3764247 (16:g.810039 A>C), rs3764246 (16:g.810143 A>G), rs2235503 (16:g.810593 C>A), rs2235504 (16:g.810655 A>G). Since a strong LD (r^2=0.94) was present between rs3764246 and rs2235504, we chose rs3764247, rs3764246 and rs2235503 for the genotyping analyses in association with SMRP.

Genotyping results in association with SMRP levels in healthy, BRD, and MPM subjects

As expected, the group of MPM patients showed a mean level of serum SMRP of 3.58 nM (±0.49, standard error of the mean SEM), significantly elevated (ANOVA, P<0.0001) when compared to the groups of healthy (0.94±0.03) or BRD (1.04±0.03) subjects. When the SMRP levels were analyzed in relation to genotypes for each SNP separately, a significant association (overall P-values calculated with ANOVA <0.0001) was found between SMRP and all the SNPs in the non-MPM category (healthy and BRD subjects). As it can be seen in Table 3 and in Figure 1, for each SNP there is an increasing and statistically significant trend of SMRP levels in relation to the carried number of variant alleles. This trend was observed among healthy individuals as well as for BRD subjects, although the comparison between heterozygotes and variant homozygotes was not significant for rs3764247 and rs2235503 in

the latter group. Interestingly, similar trends were also observed in the group of MPM patients, however no statistically significant differences were achieved for any of the SNPs (P = 0.166, 0.363 and 0.373 for rs3764247, rs3764246, and rs2235503, respectively).

In order to ascertain the global role of these SNPs, we used the three-SNPs classifier assigning the H or L genotype for each volunteer of this study. Then, a mANOVA was employed with "health status" and "classifier" as independent factors and this model confirmed that SMRP levels were associated with the promoter genotype (L vs H, P=0.001) and diagnosis (non-MPM vs MPM P <0.0001). Moreover, the interaction between these factors was not statistically significant (P=0.3730), given that also among MPM patients the group carrying the L genotype showed an average SMRP lower than the patients carrying the H genotype (however, the difference between H and L genotype within MPM patients was not statistically significant).

When SMRP was evaluated as a biomarker regardless of the genotype information, the ROC curves showed an AUC of 0.867 (95% CI = 0.841-0.890). The Youden's J index (0.566) pointed at the SMRP cut-off value of 1.28 nM, resulting in a sensitivity of 76.7% and a specificity of 79.9%. At a cut-off value of 1 nM (as suggested in previous works (Cristaudo et al., 2011; Cristaudo et al., 2010)), the sensitivity rose to 87.7%, but the specificity dropped to 64.1%. When considering the genotypes, non-MPM subjects were stratified by L or H promoter status. On the other hand, MPM patients were considered as a whole. In fact, their SMRP levels did not associate with genotypes in a statistically significant way and their stratification could have led to a reduction of the statistical power of the analysis. In Supplementary Figure 1 the distributions of SMRP values of MPM patients and controls with either H- or L-promoter are reported. In the ROC curves, the lowest rates of false positives were obtained among non-MPM subjects carrying the L-promoter, where Youden's J index rose to 0.690 (at 1.11 nM), the AUC to 0.922, and the sensitivity and specificity to 83.6% and

85.5%, respectively. ROC curves calculated for non-MPM individuals with H-promoter showed a worse performance of SMRP, with AUC of 0.801 and a decrease of specificity to 67% in correspondence of Youden's J index (1.28 nM). Figure 2 A-B-C reports the discussed ROC curves, whereas Table 4 reports the punctual values of sensitivities and specificities for each group. The different cut-off values with their corresponding sensitivity and specificity for L and H groups are reported in Supplementary Table 2 and 3, for brevity.

In vitro study on the SNPs located within the _MSLN_ promoter reported a functional role for rs3764247

In order to elucidate the biological role of the SNPs found to be associated with SMRP, an *in vitro* study was performed cloning the putative promoter region of *MSLN* upstream to a reporter gene. We then applied site-directed mutagenesis to obtain the most common haplotypes present in the population (i.e. pGL3_HAP1 A-A-C-A; pGL3_HAP2 C-G-A-G; pGL3_HAP3 A-G-C-G; pGL3_HAP4 C-A-C-A). Since the strongest association between genotype and SMRP was found among non-MPM individuals, we employed Met-5A cells as a model of non-MPM tissue. The vectors were transfected into these cells and the reporter activity under the control of promoters bearing different genetics variants was evaluated. A significant difference in RLU (overall P-value calculated with ANOVA <0.0001) was found among the constructs. When compared to pGL3_HAP1 (artificially set at 100%, ± 4% SEM), RLU values of pGL3_HAP2, pGL3_HAP3 and pGL3_HAP4 were 121% (±8%), 97% (±12%) and 182% (±18%), respectively. The pairwise comparisons revealed that the difference between pGL3_HAP1 and pGL3_HAP4 was statistically significant, as shown in Figure 3, whereas that between pGL3_HAP1 and pGL3_HAP2 is close to the statistical significance (P=0.064).

Functional annotation of the SNPs of interest

According to luciferase assay results, rs3764247 seemed to play a direct role in the regulation of the *MSLN* gene. HaploReg v4 showed that this SNP is located in DNase I hypersensitive sites (DHSs) in neuronal progenitors and astrocyte primary cells. According to RegulomeDB, it lies within enhancer regions in lung tissues and it is suggested to affect binding sites for two transcription factors, namely Staf and ZNF143. As pGL3_HAP3 did not show any difference in luciferase activity when compared to pGL3_HAP1, we could conclude that rs3764246 and rs2235504 did not exert a direct functional role in the regulation of *MSLN*. Nonetheless, rs2235505, located in the second intron of the *MSLN* gene, shows a very high LD with these SNPs (r^2>0.9). Thus, we analyzed the functional annotation available about rs2235505 in HaploReg and RegulomeDB database. It was reported to be located in DHSs in HeLa and HepG2 cell lines and to affect several transcription factor binding motifs such as BHLHE40, CTCF, PLAG1 and Rad21. It was also shown to bind RCOR1 chromatin binding protein in HeLa cells. No alteration in the splicing mechanism was predicted by SpliceAid software. A visual summary of the results of the functional study on the *MSLN* promoter is reported in Figure 4.

DISCUSSION

MSLN is a membrane glycoprotein described as functionally involved in many malignancies, including MPM. It has been repeatedly reported that the measurement of the levels of its soluble form (SMRP) could help to discriminate the MPM from the non-MPM subjects, although its performance is limited by high rates of false positives and false negatives (Cui et al., 2014). Regulatory SNPs within promoters play an important role in various diseases, including cancer (Saeed et al., 2013; Wu et al., 2014; Cingeetham et al., 2015), myocardial infarction (Domingues-Montanari et al., 2008) and diabetes (Singh et al., 2013). In the present study, we were aimed to broaden the knowledge about the biological role played by genetic variants located within *MSLN* promoter region with potential impact also on the

performance of SMRP as diagnostic biomarker. Thus, we selected four SNPs (rs3764247 A>C, rs3764246 A>G, rs2235503 C>A, rs2235504 A>G) within 1000 bp upstream the *MSLN* TSS and, in the first part of the study, we investigated the association between SMRP and genetic variants in over 700 individuals, awarding reliability against possible chance findings. We found associations between genotypes and SMRP levels among non-MPM individuals, in agreement with those reported in the cis-eQTL database within GTex portal. The genotype, together with other confounders (Park et al., 2010; Filiberti et al., 2013), contributes to the wide inter-individual variations commonly found in serum SMRP levels (Cui et al., 2014). Considering the global effect of these SNPs (summarized in the L/H classifier), different sensitivities and specificities were found when SMRP was employed as a biomarker. The inclusion of the genotype in the calculation of ROC curves led to an improved diagnostic performance, with the lowest rate of false positives in individuals carrying the L genotype, implying that high levels of SMRP could be more alarming for people carrying this genotype. In the second part of the study, we found that the genotype-dependent levels of SMRP paralleled, at least partially, the results obtained *in vitro* in non-MPM Met-5A cells, where the functional role of naturally occurring haplotypes was evaluated. Typically, the functional study of SNPs within promoters is very challenging, especially when haplotypes are studied. Since reporter vectors focus on a narrow window of the genome and SNPs could interact with each other in a complex way, the behavior of haplotype-bearing constructs is not easy to interpret, as suggested by several previous works (Terry et al., 2000; Bellini et al., 2010; Zhao et al., 2015; Shin et al., 2015). Thus, differences among haplotypes are not easily interpretable, such as those occurring between pGL3_HAP2 and pGL3_HAP4 in the present study. However, a direct effect of rs3764247 was suggested by the higher expression of pGL3_HAP4 when compared to pGL3_HAP1, and further studies are needed in order to ascertain its role in *MSLN* regulation. For instance, it is

reported to affect the binding sites of transcription factors such as Staf or ZNF143, thus future research could be directed towards the experimental validation of this interaction in mesothelial cells. Moreover, according to the luciferase assay, rs3764246 and rs2235504 are unlikely to play a direct role in *MSLN* regulation, as suggested by the similar expression of pGL3_HAP3 and pGL3_HAP1. However, rs2235505, which is located within intron 2 of *MSLN*, is in strong LD with them and it could be responsible for the differential levels of SMRP found in our association study. This SNP is also included in the list of associated SNPs in cis-eQTL database and functional annotations reported several transcription factor-binding sites affected by its variant allele. Thus, rs2235505 could be worth of further investigations including an *in vitro* functional study with a similar approach to the one performed here. Interestingly, our results are reminiscent of previous observations concerning SNPs lying within the *PSA* (prostate-specific antigen) gene promoter (Cramer et al., 2008). In fact, these SNPs were shown to contribute to individual differences among healthy men in the levels of serum PSA, a common biomarker for prostate cancer (Cramer et al., 2008). This reinforces the notion of implementing the genetic information when considering specific biomarkers in surveillance programs. Interestingly, we noticed that, similarly to what observed in non-MPM volunteers, among MPM patients rare homozygotes had the highest average levels of SMRP, whereas heterozygotes showed intermediate levels. However, these trends, as well as the difference between H and L genotypes, were not statistically significant. We hypothesize that the lack of statistical significance has to be ascribed to the relative small number of MPM patients recruited in this study. We could not collect more patients, being MPM a rare disease, however it is likely that also among patients the increase of SMRP could be more evident among carriers of the H genotype.

In conclusion, we reported that SMRP levels are affected by genetic variants, with the consequence of suggesting different “warning” thresholds for healthy subjects carrying

different genotypes. A challenging aspect of the biomarker study would be the identification of SNPs explaining the presence of false negative results, i.e. low SMRP levels among MPM patients. The recruitment of a larger sample set of MPM individuals should be required for this purpose. In the present work, a functional role for some of these SNPs was suggested and needs further investigation. These analyses could help in understanding the biological mechanisms of transcriptional regulation of *MSLN* gene and eventually contribute to explaining the high levels of this protein in MPM, shedding some lights also in the mechanisms of pleural carcinogenesis.

ACKNOWLEDGMENTS

This work was supported by GIME (Gruppo Italiano Mesotelioma) onlus and by Ministero della Salute-Bando Ricerca Finalizzata 2009 (RF-2009-1529895). There are no conflicts of interest with regard to this manuscript. All authors are aware of and agree to the content of the paper and their being listed as an author on the paper.

REFERENCES

1. Bellini I, Pitto L, Marini MG, Porcu L, Moi P, Garritano S, Boldrini L, Rainaldi G, Fontanini G, Chiarugi M, Barale R, Gemignani F, et al. 2010. DeltaN133p53 expression levels in relation to haplotypes of the TP53 internal promoter region. Hum Mutat 31:456-465.
2. Bera TK and Pastan I. 2000. Mesothelin is not required for normal mouse development or reproduction. Mol Cell Biol 20:2902-2906.
3. Cingeetham A, Vuree S, Jiwatani S, Kagita S, Dunna NR, Meka PB, Gorre M, Annamaneni S, Digumarti R, Sinha S, Satti V. 2015. Role of the MDM2 promoter polymorphism (-309T>G) in acute myeloid leukemia development. Asian Pac J Cancer Prev 16:2707-2012.
4. Cramer SD, Sun J, Zheng SL, Xu J, Peehl DM. 2008. Association of prostate-specific antigen promoter genotype with clinical and histopathologic features of prostate cancer. Cancer Epidemiol Biomarkers Prev 17:2451-2457.
5. Cristaudo A, Foddis R, Bonotti A, Simonini S, Vivaldi A, Guglielmi G, Bruno R, Gemignani F, Landi S. 2011. Two novel polymorphisms in 5' flanking region of the mesothelin gene are associated with soluble mesothelin-related peptide (SMRP) levels. Int J Biol Markers 26:117-123.
6. Cristaudo A, Foddis R, Bonotti A, Simonini S, Vivaldi A, Guglielmi G, Bruno R, Landi D, Gemignani F, Landi S. 2010. Polymorphisms in the putative micro-RNA-binding sites of mesothelin gene are associated with serum levels of mesothelin-related protein. Occup Environ Med 67:233-236.
7. Cristaudo A, Foddis R, Vivaldi A, Guglielmi G, Dipalma N, Filiberti R, Neri M, Ceppi M, Paganuzzi M, Ivaldi GP, Mencoboni M, Canessa PA, et al. 2007. Clinical significance

of serum mesothelin in patients with mesothelioma and lung cancer. Clin Cancer Res 13:5076-5081.

8. Cui A, Jin XG, Zhai K, Tong ZH, Shi HZ. 2014. Diagnostic values of soluble mesothelin-related peptides for malignant pleural mesothelioma: updated meta-analysis. BMJ Open 4:e004145.
9. Domingues-Montanari S, Subirana I, Tomás M, Marrugat J, Sentí M. 2008. Association between ESR2 genetic variants and risk of myocardial infarction. Clin Chem 54:1183-1189.
10. Filiberti R, Marroni P, Mencoboni M, Mortara V, Caruso P, Cioè A, Michelazzi L, Merlo DF, Bruzzone A, Bobbio B, Del Corso L, Galli R, et al. 2013. Individual predictors of increased serum mesothelin in asbestos-exposed workers. Med Oncol 30:422.
11. Fukuoka K, Kuribayashi K, Yamada S, Tamura K, Tabata C, Nakano T. 2013. Combined serum mesothelin and carcinoembryonic antigen measurement in the diagnosis of malignant mesothelioma. Mol Clin Oncol 1:942-948.
12. Garritano S, De Santi C, Silvestri R, Melaiu O, Cipollini M, Barone E, Lucchi M, Barale R, Mutti L, Gemignani F, Bonotti A, Foddis R, et al. 2014. A common polymorphism within MSLN affects miR-611 binding site and soluble mesothelin levels in healthy people. J Thorac Oncol 9:1662-1168.
13. Hassan R, Bera T, Pastan I. 2004. Mesothelin: a new target for immunotherapy. Clin Cancer Res 10:3937-3942.
14. Hassan R, Cohen SJ, Phillips M, Pastan I, Sharon E, Kelly RJ, Schweizer C, Weil S, Laheru D. 2010. Phase I clinical trial of the chimeric anti-mesothelin monoclonal antibody MORAb-009 in patients with mesothelin-expressing cancers. Clin Cancer Res 16:6132-6138.

15. Hassan R and Ho M. 2008. Mesothelin targeted cancer immunotherapy. Eur J Cancer 44:46-53.
16. Hassan R, Sharon E, Thomas A, Zhang J, Ling A, Miettinen M, Kreitman RJ, Steinberg SM, Hollevoet K, Pastan I. 2014. Phase 1 study of the antimesothelin immunotoxin SS1P in combination with pemetrexed and cisplatin for front-line therapy of pleural mesothelioma and correlation of tumor response with serum mesothelin, megakaryocyte potentiating factor, and cancer antigen 125. Cancer 120:3311-3319.
17. Hollevoet K, Bernard D, De Geeter F, Walgraeve N, Van den Eeckhaut A, Vanholder R, Van de Wiele C, Stove V, van Meerbeeck JP, Delanghe JR. 2009. Glomerular filtration rate is a confounder for the measurement of soluble mesothelin in serum. Clin Chem 55:1431-1433.
18. Ke Y, Reddel RR, Gerwin BI, Reddel HK, Somers AN, McMenamin MG, LaVeck MA, Stahel RA, Lechner JF, Harris CC. 1989. Establishment of a human in vitro mesothelial cell model system for investigating mechanisms of asbestos-induced mesothelioma. Am J Pathol 134:979-991.
19. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, et al. 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580-585.
20. Panou V, Vyberg M, Weinreich UM, Meristoudis C, Falkmer UG, Røe OD. 2015. The established and future biomarkers of malignant pleural mesothelioma. Cancer Treat Rev 41:486-495.
21. Park EK, Thomas PS, Creaney J, Johnson AR, Robinson BW, Yates DH. 2010. Factors affecting soluble mesothelin related protein levels in an asbestos-exposed population. Clin Chem Lab Med 48:869-874.

22. Pass HI, Wali A, Tang N, Ivanova A, Ivanov S, Harbut M, Carbone M, Allard J. 2008. Soluble mesothelin-related peptide level elevation in mesothelioma serum and pleural effusions. Ann Thorac Surg 85:265-272.
23. Robinson BW, Creaney J, Lake R, Nowak A, Musk AW, de Klerk N, Winzell P, Hellstrom KE, Hellstrom I. 2003. Mesothelin-family proteins and diagnosis of mesothelioma. Lancet 362:1612-1616.
24. Saeed HM, Alanazi MS, Alshahrani O, Parine NR, Alabdulkarim HA, Shalaby MA. 2013. Matrix metalloproteinase-2 C(-1306)T promoter polymorphism and breast cancer risk in the Saudi population. Acta Biochim Pol 60:405-409.
25. Shin HJ, Kim JY, Cheong HS, Na HS, Shin HD, Chung MW. 2015. Functional Study of Haplotypes in UGT1A1 Promoter to Find a Novel Genetic Variant Leading to Reduced Gene Expression. Ther Drug Monit 37:369-374.
26. Singh K, Agrawal NK, Gupta SK, Singh K. 2013. A functional single nucleotide polymorphism -1562C>T in the matrix metalloproteinase-9 promoter is associated with type 2 diabetes and diabetic foot ulcers. Int J Low Extrem Wounds 12:199-204.
27. Terry CF, Loukaci V, Green FR. 2000. Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. J Biol Chem 275:18138-18144.
28. Wu H, Zhang K, Gong P, Qiao F, Wang L, Cui H, Sui X1, Gao J, Fan H. 2014. A novel functional TagSNP Rs7560488 in the DNMT3A1 promoter is associated with susceptibility to gastric cancer by modulating promoter activity. PLoS One 9:e92911.
29. Zhao N, Xiao J, Zheng Z, Fei G, Zhang F, Jin L, Zhong C. 2015. Single-nucleotide polymorphisms and haplotypes of non-coding area in the CP gene are correlated with Parkinson's disease. Neurosci Bull 31:245-256.

FIGURE LEGENDS

Figure 1. Association between genetic variants within the *MSLN* promoter (i.e. rs3764247, rs3764246 and rs2235503) and SMRP levels in healthy (A), BRD (B) and MPM (C) subjects. Asterisks show a statistical significance ($P < 0.05$) in the Tukey's test for pairwise differences within the ANOVA model. The columns represent mean values, the bars show standard error of the mean (SEM).

Figure 2. ROC curves obtained with MedCalc software comparing (A) the whole non-MPM group vs MPM; (B) L genotype group (non-MPM subjects carrying common homozygote genotype for all the SNPs) vs MPM; (C) H genotype group (all the other non-MPM individuals) vs MPM. AUC, sensitivities, specificities and cut-off value ("criterion") at the Youden's J index are shown in the figure.

Figure 3. Luciferase assay results on Met-5A cells when the four haplotypes-mimicking plasmids were co-transfected with pRL-SV40 control vector. RLU obtained with pGL3_HAP1 transfection is reported as 100% and used as reference for statistical evaluation. The columns represent mean values, the bars show standard error of the mean (SEM).

Figure 4. Summary of the results of the functional study on the MSLN promoter. The grey square represents the cloned region of about 1000 bp with the SNPs having a MAF>0.05. The transcription factors binding to the polymorphic sites and affected by these SNPs (according to HaploReg and/or Regulome DB) are also reported. The plot linking the SNPs each other shows the r^2 values of LD as a greyscale (plotted with Haploview). The four most common haplotypes and their corresponding differences in luciferase activity are shown in the lower part of the figure, where a qualitative trend of RLU is reported for each haplotype-mimicking plasmid, referred to HAP_1 as reference.