TITLE PAGE:

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

Authors: Chiara De Santi, PhD,[1,§] Perla Pucci, MS,[1] Alessandra Bonotti, PhD,[2] Ombretta Melaiu, PhD,[1,ǂ] Monica Cipollini, PhD,[1] Roberto Silvestri, BS,[1] Veronika Vymetalkova, PhD,[3] Elisa Barone, MS,[1], Elisa Paolicchi, PhD,[1] Alda Corrado, MS,[1] Irene Lepori, MS,[1] Irene Dell'Anno, MS,[1] Lucia Pellè, MS,[1] Pavel Vodicka, PhD[3] Luciano Mutti, Associate Professor,[4] Rudy Foddis, MD,[5] Alfonso Cristaudo, Associate Professor,[5] Federica Gemignani, Associate Professor,[1] and Stefano Landi, Full Professor,[1,*].

Affiliation:

1. Department of Biology, University of Pisa, Via Derna 1, Pisa, Italy.
2. Preventive and Occupational Medicine, University Hospital of Pisa, Pisa, Italy.
3. Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic.
4. School of Environment and Life Sciences, University of Salford, Manchester, United Kingdom.
5. Department of Translational Research and of new Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy.

[§]Present address: Respiratory Research Division, Department of Medicine, Education and Research Centre, Royal College of Surgeons in Ireland, Beaumont Hospital, Dublin 9, Ireland.

[ǂ]Present address: Immuno-Oncology Laboratory, Department of Paediatric Haematology/Oncology, Ospedale Pediatrico Bambino Gesù, Viale di S. Paolo 15, 00146 Rome, 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.

Word count: 3485 words

ABSTRACT AND KEYWORDS

Background: Soluble mesothelin-related peptide (SMRP) is a promising diagnostic biomarker for malignant pleural mesothelioma (MPM), but various confounders hinder its usefulness in surveillance programmes. We previously showed that a single nucleotide polymorphism (SNP) within the 3'untranslated region (3'UTR) of the mesothelin (MSLN) gene could affect the levels of SMRP.

Objectives To focus on SNPs located within MSLN promoter as possible critical genetic variables in determining SMRP levels.

Methods: The association between SMRP and SNPs was tested in 689 non-MPM subjects and 70 patients with MPM. Reporter plasmids carrying the four most common haplotypes were compared in a dual luciferase assay, and in silico analyses were performed to investigate the putative biological role of the SNPs.

Results: We found a strong association between serum SMRP and variant alleles of rs3764247, rs3764246 (in strong linkage disequilibrium with rs2235504) and rs2235503 in non-MPM subjects. Inclusion of the genotype information led to an increase in SMRP specificity from 79.9% to 85.5%. Although not statistically significant, the group with MPM showed the same trend of association. According to the in vitro luciferase study, rs3764247 itself had a functional role. In silico approaches showed that the binding sites for transcription factors such as Staf and ZNF143 could be affected by this SNP. The other SNPs were shown to interact with each other in a more complex way.

Conclusions: These data support the suggestion that SMRP performance is affected by individual (ie, genetic) variables and that MSLN expression is influenced by SNPs located within the promoter regulatory region.

KEY WORDS: mesothelioma, polymorphisms, health surveillance

INTRODUCTION

Mesothelin (MSLN) is a membrane-bound glycoprotein physiologically expressed by the mesothelial tissues of pleura, peritoneum and pericardium.1 Although its biological function is still unknown,2 many types of cancer, including malignant pleural mesothelioma (MPM), show increased expression of MSLN compared with their non-malignant counterparts.3 MPM is a highly aggressive tumour of the pleural cavities, associated with asbestos exposure and characterised by challenging diagnosis and poor prognosis.4 In recent years, several research groups have suggested that MSLN might be helpful in the management of MPM, both as a diagnostic tool5 6 and as a putative therapeutic target.7 8 In particular, high levels of the soluble form of MSLN, the so-called SMRP (soluble mesothelin-related peptides), were repeatedly found in serum samples of patients with MPM in comparison with various types of control groups.6 9 10 Nonetheless, in spite of the initial findings, the real usefulness of SMRP within surveillance programmes is hindered by a relatively high rate of false-negative and false-positive results.11 Various demographic and clinical variables were reported as possible confounders, such as body mass index, age, glomerular filtration rate and lung function.12–14 Genetic factors were also shown to affect SMRP levels in non-MPM subjects. Thus, the inclusion of individuals' genetic information could improve the receiver operating characteristic (ROC) curves, leading to slight improvement of the performance of SMRP as biomarker.15

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.15 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.16 Healthy subjects carrying the variant allele of rs3764247 A>C (reported as New1 in the original publication) showed increased SMRP levels compared with those carrying the AA genotype.16 This could be ascribed to a different regulatory pattern depending on the presence of the variant or common allele. In this present study we analysed a large sample set and 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. These findings could help to refine the use of SMRP as diagnostic biomarker and shed some light on the regulatory mechanisms of the *MSLN* gene.

MATERIALS AND METHODS

Selection of SNPs

In the pilot study, an association between rs3764247 and SMRP was found.16 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) they must lie within 1000 bp (arbitrarily chosen) upstream of the *MSLN* transcriptional start site (TSS); (ii) the frequency of the rare allele must exceed >0.05; (iii) they must be reported as associated with *MSLN* mRNA expression in 278 lung tissue samples according to GTex portal (http://www. gtexportal. org/ home/).17 The linkage disequilibrium (LD) between the selected SNPs (ie, 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 (Northern Europeans from Utah) 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 programme on workers previously exposed to asbestos, as described in detail by Garritano *et al*.15 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 provided written informed consent. For genotyping, whole blood and serum samples were obtained by venepuncture 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 (ie, 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 immunosorbent assay according to the manufacturer's instructions (Mesomark, Fujirebio Diagnostics, Japan).

	Healthy	**BRD**	**MPM**
Sex			
Male	97.8% (363)	98.4% (313)	85.7% (60)
Female	2.2% (8)	1.6% (5)	14.3% (10)
Age (years)			
Average	58.9±13.5	65.8±10.7	70.5±10.8
Median	60	66	71
Smoking			
Smokers	16.2% (60)	16.2% (51)	8.3% (6)
Ex-Smokers	39.9% (148)	54.0% (172)	58.3% (41)
Non-Smokers	43.9% (163)	29.8% (95)	33.4% (23)
Asbestos exposure (years)			
Average	15.6±11.5	20.0±10.7	23.2±19.1
Median	15	20	25
Diagnosis			
		*Pleural 56.5% (180)	eMPM 60.0% (42)
		**Lung 35% (111)	sMPM 21.4% (15)
		***Airways 8.5% (27)	bMPM 18.6% (13)
MAF			
rs3764247 (C)	0.20	0.18	0.20
rs3764246 (G)	0.24	0.22	0.25
rs2235503 (A)	0.15	0.13	0.16
Total	371	318	70

Legend: *Pleural diseases include: pleural plaques 30.8% (98), pleural thickening 20.1% (64), pleural effusion 1.9% (6), pleuritis 3.7% (12). **Lung diseases include emphysema 10.1% (32), lung fibrosis 4.1% (13), nodules 15.1% (48), asbestosis 5.7% (18). ***Airways diseases correspond to bronchiectasis 8.5% (27). BRD Benign Respiratory Disease. eMPM Epithelioid Malignant Pleural Mesothelioma; sMPM Sarcomatoid Malignant Pleural Mesothelioma; bMPM Biphasic Mesothelioma. MAF= minor allele frequency; the minor allele for each SNP is reported in brackets.

Association analyses between genotypes and SMRP levels

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 + patients with BRD) 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 (ie, 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 GraphPad PRISM 4.0 (San Diego, California, USA).

ROC curves were generated with MedCalc statistical software (version 12.7.2.0, MedCalc Software, Belgium) comparing the non-MPM and MPM groups. 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 the L (n=374) or H (n=315) genotype, versus the whole group of patients with MPM (n=70). In a second analysis, the group with MPM was also stratified by L (n=37) and H (n=33) genotype and the ROC curves were repeated.

Construction of plasmids

The putative human MSLN promoter from nucleotides −1 to−1073 relative to the TSS of RefSeq NM_005823.5 was amplified by Q5 high-fidelity DNA polymerase (NEB, Ipswich, USA).

As template, an individual carrying the common homozygote genotype for all SNPs in the study was selected from our sample set. The resultant PCR amplicon was subsequently cloned into the XhoI site

of the pGL3-basic vector (Promega, Madison, Wisconsin, USA) using CloneEZ PCR Cloning Kit (GenScript, Piscataway, USA). This construct, bearing the most common haplotype in the TSI population (ie, common allele for every SNP), 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, California, 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 is reported in online supplementary table 1.

Cell culture and luciferase reporter assays

Non-malignant transformed human pleural mesothelial cells (Met-5A)18 were purchased from ATCC (American Type Culture Collection) and cultured in Medium 199 (Gibco in Life Technologies, Monza, Italy) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 3 nM epidermal growth factor, 400 nM hydrocortisone and 870 nM insulin. Human epithelioid malignant mesothelioma cells (Mero-14) were kindly donated by Istituto Tumori of Genova (National Research Council, Genova, Italy) and maintained in Dulbecco's modified Eagle's medium (Lonza, Maryland, USA). Met-5A and Mero-14 cells were maintained in a humidified incubator at 37°C in 5% CO2.

In three independent experiments, Met-5A and Mero-14 cells were seeded in 24-well plates at a final density of 50 000 cells/well and incubated for 24 hours. Cells were then co-transfected at 60–80% confluence with 400 ng of pGL3_HAP1/2/3/4 and 10 ng of the Renilla pRL-SV40 internal control vector (Promega, Madison, USA) using Attractene reagent (Qiagen, Hilden, Germany). Twenty-four hours after transfection, a dual-luciferase reporter assay (Promega, Madison, USA) was performed.

Relative luciferase units (RLU) were expressed as mean value of the firefly luciferase/Renilla luciferase ratio of three independent experiments. One-way ANOVA was performed on the RLU values to assess statistically significant differences among the four transfected plasmids. Dunnett's multiple comparison tests were performed to assess pairwise differences between the variant vectors

compared with pGL3_HAP1 plasmid, which carried the most common haplotype (ie, common allele for every SNP). Two-way ANOVA was also performed to compare RLU values of the four different plasmids among the two cell lines.

RESULTS

SNPs selection

In order to identify the genetic variants within MSLN proximal promoter (~1000 bp upstream from 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 4.9×10–6 to 4.4×10–33. The region spanning MSLN TSS showed the highest associated SNPs. Table 2 lists the top 10 associated SNPs (ie, the SNPs with the most significant p value according to GTex) with their main features. Among the 86 associated SNPs, we selected those located within the 1000 bp upstream from the TSS—that is, 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 (r2=0.94) was present between rs3764246 and rs2235504, we chose rs3764247, rs3764246 and rs2235503 for the genotyping analyses in association with SMRP.

SNP	P-value from Gtex	Effect size	Location	Other SNPs in LD
rs2235503	$4.4x10^{-33}$	0.98	Promoter (-171 bp from TSS)	rs12597489 (r^2= 0.8)
rs2235505	$4.1x10^{-28}$	0.84	Intron 2	rs3764246 (r^2=0.9) rs2235504 (r^2=0.92) rs9925870 (r^2=0.85)
rs2235504	$3.6x10^{-27}$	0.81	Promoter (-109 bp a from TSS)	rs3764246 (r^2= 0.94) rs2235505 (r^2= 0.92)
rs3764246	$7.5x10^{-27}$	0.80	Promoter (-621 bp from TSS)	rs2235504 (r^2=0.94) rs2235505 (r^2=0.9)
rs12600012	$5.6x10^{-25}$	0.96	Intron 2	rs12597489 (r^2= 0.85) rs3765319 (r^2= 0.81)
rs3764247	$7.5x10^{-25}$	0.79	Promoter (-724 bp from TSS)	-
rs9925870	$2.9x10^{-24}$	0.79	Intron 2	rs2235505 (r^2= 0.85)
rs7198927	$1.3x10^{-23}$	0.94	Promoter (-2593 bp from TSS)	rs67623411 (r^2= 0.89) rs12597489 (r^2= 0.88)
rs12597489	$1.5x10^{-23}$	0.94	Promoter (-1602 bp from TSS)	rs67623411 (r^2= 0.86) rs7198927 (r^2= 0.88) rs2235503 (r^2= 0.8) rs12600012 (r^2= 0.85)
rs7185523	$8.4x10^{-22}$	0.77	Intron 6	rs7185150 (r^2= 0.89)

Table 2. Top ten SNPs in association with *MSLN* mRNA expression in 278 lung tissues according to GTex portal. For each SNP, the P-value of the association and the effect size (i.e. the effect of the alternative allele relative to the reference allele on the mRNA expression) according to GTex is reported, together with its position relative to *MSLN* gene and all the SNPs with an $r^2 \leq 0.8$. TSS=transcriptional start site; NM_005823.5 was used as reference.

Genotyping results in association with SMRP levels in healthy subjects patients with BRD MPM

As expected, the group of patients with MPM showed a mean serum level of SMRP of 3.58 nM (±0.49, SEM), significantly higher than for healthy subjects (0.94±0.03) or patients with BRD (1.04±0.03) (ANOVA, $p<0.0001$). When the SMRP levels were analysed 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 subjects and patients with BRD). As 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 number of variant alleles carried. This trend was observed among healthy individuals and patients with BRD, although the comparison between heterozygotes and variant homozygotes was not significant for rs3764247 and rs2235503 in the latter group according to Tukey's multiple comparison test. Interestingly, similar trends were also seen in the group of patients with MPM; 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), probably owing to low statistical power when comparing subgroups of patients with MPM. 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. 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 ($p=0.373$), given that among patients with MPM the group carrying the L genotype also showed an average SMRP lower than the patients carrying the H genotype (however, the difference between H and L genotype within patients with MPM was not statistically significant). When SMRP was evaluated as a biomarker regardless of the genotype information, the ROC curves showed an area under the curve (AUC) of 0.867 (95% CI 0.841 to 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),16 19 the sensitivity rose to 87.7%, but the specificity dropped to 64.1%. When considering the genotypes, in a first analysis, non-MPM subjects were stratified by L (n=374) or H (n=315) promoter status, whereas patients with MPM were considered as a whole. In fact, their SMRP levels did not associate with genotypes in a statistically significant way and their stratification might have led to a reduction of the statistical power of the analysis. The distributions of these SMRP values are reported in online supplementary figure 1. In the ROC curves, the lowest rates of false-positive results 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 the H promoter showed a worse performance, with AUC of 0.801 and a decrease of specificity to 67% corresponding to Youden's J index (1.28 nM).

Online supplementary figure 2 A-B-C reports these ROC curves, whereas online supplementary table 2 reports the 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 online supplementary tables 3 and 4. A second analysis was attempted by stratifying both non-MPM and MPM volunteers for L and H genotypes, despite the small number of patients with MPM falling in

the two genotypic groups (37 and 33, respectively). In the ROC curves, the lowest rates of false-positive results were again obtained among subjects carrying the L promoter, where Youden's J index pointed at an optimal cut-off value of 1.11 nM, the AUC to 0.914 and the sensitivity and specificity to 79.5% and 85.5%, respectively. ROC curves calculated for individuals with the H promoter showed a worse performance for SMRP, with AUC of 0.829 and a decrease of specificity to 67%, in correspondence with Youden's J index (1.28 nM). These ROC curves are reported in online supplementary figure 2B,C (right panel) in parallel with the curves obtained without stratifying the MPM group according to the promoter genotype (left panel).

SNP	Healthy subjects (371)	BRD subjects (318)	MPM subjects (70)
rs3764247	AA 0.76±0.03 (237, 64%) AC 1.20±0.05 (119, 32%) CC 1.75±0.13 (15, 4%)	AA 0.92±0.04 (213, 67%) AC 1.29±0.06 (95, 30%) CC 1.62±0.19 (10, 3%)	AA 3.33±0.60 (46, 66%) AC 3.17±1.10 (18, 26%) CC 6.78±1.74 (6, 8%)
rs3764246	AA 0.73±0.03 (214, 58%) AG 1.11±0.04 (136, 37%) GG 1.86±0.10 (21, 5%)	AA 0.88±0.04 (193, 61%) AG 1.25±0.06 (109, 34%) GG 1.80±0.14 (16, 5%)	AA 3.14±0.66 (39, 56%) AG 3.77±0.82 (26, 37%) GG 5.97±1.91 (5, 7%)
rs2235503	CC 0.76±0.03 (268, 72%) CA 1.29±0.05 (94, 25%) AA 2.20±0.15 (9, 3%)	CC 0.91±0.04 (240, 75%) CA 1.45±0.07 (72, 23%) AA 1.82±0.24 (6, 2%)	CC 3.23±0.59 (51, 73%) CA 4.05±1.07 (15, 21%) AA 6.15±2.13 (4, 6%)

Table 3. SMRP values for each genotype in each health-status group. The values are reported as mean (nM) ± standard error (SEM). Absolute numbers and percentages are given in brackets.

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* (−1 to −1073 relative to the TSS) upstream from a reporter gene. We then applied site-directed mutagenesis to obtain the most common haplotypes present in the population, since we reasoned that the functional role of SNPs should be investigated in a genetic environment that allows SNP-to-SNP interactions as they are likely to happen physiologically. The four most common haplotypes in the TSI/CEU population according to Haploview are rs3764247(A)–rs3764246(A)–rs2235503(C)–rs2235504(A) (HAP1, 71%) (common allele for all SNPs), rs3764247(C)–rs3764246(G)–rs2235503(A)–rs2235504(G) (HAP2, 15%) (variant allele for each SNP), rs3764247(A)–rs3764246(G)–rs2235503(C)–rs2235504(G) (HAP3,

8%) (variant allele for second and fourth SNP), rs3764247(C)–rs3764246(A)–rs2235503(C)–rs2235504(A) (HAP4, 4%) (variant allele for the first SNP). Therefore, we obtained four plasmids (pGL3_HAP1/2/3/4) carrying the above SNP sequence and used a luciferase assay to study the activity of the promoter. The vectors were transfected into Met-5A and Mero-14 cells and the reporter activity under the control of promoters bearing different genetics variants was evaluated.

In comparison with pGL3_HAP1 (set at 100%,±4% SEM), RLU values of pGL3_HAP2, pGL3_HAP3 and pGL3_HAP4 were 121% (±8%), 97% (±12%) and 182% (±18%) in Met-5A, respectively (figure 2A). In Mero-14 these values were 194% (±19%), 153% (±34%) and 191% (±34%), as shown in figure 2B. A significant difference in RLU (overall p value calculated with mANOVA <0.0001) was found among the constructs in both cell lines, whereas the interaction between haplotypes and cell lines was not statistically significant (p interaction=0.185), suggesting that the constructs gave similar responses in both cell lines (summarised in figure 2C). The Dunnett's pairwise comparisons showed that in both cell lines pGL3_HAP2 was higher than pGL3_HAP1 (p values of 0.064 and 0.028 in Met-5A and Mero-14, respectively), pGL3_HAP3 was not statistically different from pGL3_HAP1 (p=0.221 and 0.358), whereas pGL3_HAP4 led to the highest RLU signal (p=0.0031 and 0.034 in comparison with pGL3_HAP1). The statistically significant differences compared with pGL3_HAP1 are shown in figure 2 marked with an asterisk.

Discussion

MSLN is a membrane glycoprotein described as functionally involved in many malignancies, including MPM. It has been repeatedly reported that measurement of the levels of its soluble form (SMRP) might help to discriminate between patients with MPM and non-MPM subjects, although its performance is limited by high rates of false-positive and false-negative results.11 Regulatory SNPs within promoters play an important role in various diseases, including cancer,20–22 myocardial infarction23 and diabetes.24 In this study, we aimed at broadening our knowledge of the biological role played by genetic variants located within MSLN promoter region, with potential impact also on

the performance of SMRP as a diagnostic biomarker. Thus, we selected four SNPs (rs3764247 A>C, rs3764246 A>G, rs2235503 C>A, rs2235504 A>G) within 1000 bp upstream from the MSLN TSS and, in the first part of the study, we investigated the association between SMRP and genetic variants in over 700 individuals, assigning reliability against possible chance findings. Although post-transcriptional and post-translational regulatory mechanisms (such as alternative splicing, microRNAs and proteolytic cleavage) could impair the correlation between MSLN mRNA and its product SMRP, we found significant associations between genotypes and SMRP levels, in agreement with those reported in the cis-eQTL database within the GTex portal. The genotype, together with other confounders,13 14 contributes to the wide interindividual variations commonly found in serum SMRP levels.11 Considering the global effect of these SNPs (summarised 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-positive results in individuals carrying the L genotype, implying that high levels of SMRP could be more worrying for people carrying this genotype. In surveillance screening, information about the genotype could be helpful in interpreting the SMRP measurement, as shown by data from ROC curves, where both clinical groups were stratified by L/H 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, where the functional role of naturally occurring haplotypes was evaluated. Overall, a direct effect of rs3764247 was suggested by the higher expression of pGL3_HAP4 and pGL3_HAP2 in comparison with pGL3_HAP1; further studies are needed to ascertain its role in MSLN regulation. HaploReg v4 (www.broadinstitute.org/mammals/haploreg) showed that this SNP is located in DNase I hypersensitive sites (DHSs) in neuronal progenitors and astrocyte primary cells. According to RegulomeDB (http://www.regulomedb.org/), it lies within enhancer regions in lung tissues and it is suggested that it affects binding sites for two transcription

factors, namely Staf and ZNF143. Thus future research could be directed towards the experimental validation of these interactions 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. Because these two SNPs were found associated with in vivo SMRP levels and with MSLN mRNA (as cis-eQTL within the GTex database), a different mechanism should be evoked. A SNP in strong LD with them might be responsible for these observations. Rs2235505, located within intron 2 of the MSLN gene, has an r2=0.92 with rs2235504 and r2=0.9 with rs3764246. Moreover, this intronic SNP is listed in GTex as cis-eQTL of MSLN mRNA, and functional annotations reported its localisation in DHSs in HeLa and HepG2 cell lines and its ability 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. Thus, rs2235505 might be worth further investigation, including an in vitro study to assess whether it might be the functional SNP responsible for the in vivo observed associations. A visual summary of the results of the functional study on the MSLN promoter is reported in Supplementary Figure 3.

Beside the potential location of the SNPs in DHSs, other regulatory mechanisms could be affected by single nucleotide variations. SNPs located within CpG islands in the promoter were previously shown to affect expression of the neighbour mRNA and eventually, to be associated with the pathological condition.25 26 In MPM, a clear hypomethylation of MSLN promoter was observed by Tan and coauthors27 and interestingly, Nelson and collaborators found an association between MSLN promoter hypomethylation and high levels of SMRP in patients with MPM.28 Therefore, it is reasonable to suggest that SNPs located in the MSLN promoter within a CpG island could affect the methylation status and ultimately, the SMRP levels. Nonetheless, the SNPs analysed in this study do not fall within the region examined by the Nelson or Tan research groups, thus their role in epigenetic regulation would need to be further explored.

Interestingly, our evidence of association between genetic variants and biomarker levels is reminiscent of previous observations concerning SNPs lying within the PSA (prostate-specific antigen) gene promoter.29 These SNPs were shown to contribute to individual differences among healthy men in the levels of serum PSA, a common biomarker for prostate cancer.29 This reinforces the need to make use of the genetic information when considering specific biomarkers in surveillance programmes. Interestingly, we noticed that, as seen in non-MPM volunteers, patients with MPM rare homozygotes had the highest average levels of SMRP, whereas heterozygotes showed intermediate levels. However, these trends, and the difference between H and L genotypes, were not statistically significant. We investigated whether this could be due to an association between tumour histology and genotypes (since SMRP is more frequently elevated in the epithelioid subtype), but no relation was found (data not shown). Therefore, we hypothesised that the lack of statistical significance might be ascribed to the relatively small number of patients with MPM recruited to this study. This hypothesis was supported by a post hoc power analysis (data not shown) and by the positive results we obtained in luciferase assays in MPM Mero-14 cells. We could not gather together more patients, as MPM is a rare disease, but it is likely that among patients the increase of SMRP might be more evident among carriers of the H genotype.

In conclusion, we reported that SMRP levels are affected by genetic variants, resulting in 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—that is, low SMRP levels among patients with MPM. The recruitment of a larger sample of patients with MPM would be required for this purpose. The present work suggested that some of these SNPs have a functional role and this needs further investigation. These analyses could help in understanding the biological mechanisms of transcriptional regulation of the MSLN gene and eventually contribute to explaining the high levels of this protein in MPM, shedding some light also on the mechanisms of pleural carcinogenesis.

Acknowledgments

The authors thank Roberto Favoni (IRCCS A.O.U. San Martino-IST, Laboratory of Gene Transfer) for the donation of the Mero-14 MPM cell line. There are no conflicts of interest in regard to this paper.

AUTHOR CONTRIBUTIONS STATEMENT

The manuscript was made possible thanks to the following contributions. Stefano Landi, Federica Gemignani, Alfonso Cristaudo and Luciano Mutti conceived the idea at the basis of the study. Alessandra Bonotti, Rudy Foddis, and Alfonso Cristaudo from the Occupational Medicine of Cisanello Hospital in Pisa contributed with essential biological samples and the related clinical features. Chiara De Santi, Monica Cipollini, Elisa Barone, Elisa Paolicchi, Alda Corrado, Irene Lepori, Irene Dell'Anno and Lucia Pellè carried out the DNA extractions for the genotyping. Chiara De Santi, Monica Cipollini and Stefano Landi performed the statistical analyses on the association between genotypes and SMRP levels. Chiara De Santi and Perla Pucci obtained the plasmids via cloning and mutagenesis and performed the *in vitro* luciferase assay on Met-5A cells, with the contribution of Ombretta Melaiu. Roberto Silvestri and Veronika Vymetalkova carried out the luciferase assays on Mero-14 MPM cells. Chiara De Santi wrote and revised the paper. Federica Gemignani and Pavel Vodicka helped to evaluate, edit and revise the manuscript. Stefano Landi, Alfonso Cristaudo and Luciano Mutti supervised the development of work, managed the data interpretation and revised the manuscript.

REFERENCES

1. Hassan R, Bera T, Pastan I. Mesothelin: a new target for immunotherapy. *Clin Cancer Res* 2004;10:3937-3942.
2. Bera TK and Pastan I. Mesothelin is not required for normal mouse development or reproduction. *Mol Cell Biol* 2000;20:2902-2906.
3. Hassan R and Ho M. Mesothelin targeted cancer immunotherapy. *Eur J Cancer* 2008;44:46-53.
4. Panou V, Vyberg M, Weinreich UM, et al. The established and future biomarkers of malignant pleural mesothelioma. *Cancer Treat Rev* 2005;41:486-495. doi: 10.1016/j.ctrv.2015.05.001.
5. Robinson BW, Creaney J, Lake R, et al. Mesothelin-family proteins and diagnosis of mesothelioma. *Lancet* 2003;362:1612-1616.
6. Cristaudo A, Foddis R, Vivaldi A, et al. Clinical significance of serum mesothelin in patients with mesothelioma and lung cancer. *Clin Cancer Res* 2007;13:5076-5081.
7. Hassan R, Cohen SJ, Phillips M, et al. Phase I clinical trial of the chimeric anti-mesothelin monoclonal antibody MORAb-009 in patients with mesothelin-expressing cancers. *Clin Cancer Res* 2010;16:6132-6138. doi: 10.1158/1078-0432.CCR-10-2275.
8. Hassan R, Sharon E, Thomas A, et al. 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* 2014;120:3311-3319. doi: 10.1002/cncr.28875.
9. Pass HI, Wali A, Tang N, et al. Soluble mesothelin-related peptide level elevation in mesothelioma serum and pleural effusions. *Ann Thorac Surg* 2008;85:265-272.
10. Fukuoka K, Kuribayashi K, Yamada S, et al. Combined serum mesothelin and carcinoembryonic antigen measurement in the diagnosis of malignant mesothelioma. *Mol Clin Oncol* 2013;1:942-948.

11. Cui A, Jin XG, Zhai K, et al. Diagnostic values of soluble mesothelin-related peptides for malignant pleural mesothelioma: updated meta-analysis. *BMJ Open* 2014;4:e004145. doi: 10.1136/bmjopen-2013-004145.

12. Hollevoet K, Bernard D, De Geeter F, et al. Glomerular filtration rate is a confounder for the measurement of soluble mesothelin in serum. *Clin Chem* 2009;55:1431-1433. doi: 10.1373/clinchem.2008.121913.

13. Park EK, Thomas PS, Creaney J. Factors affecting soluble mesothelin related protein levels in an asbestos-exposed population. *Clin Chem Lab Med* 2010;48:869-874. doi: 10.1515/CCLM.2010.165.

14. Filiberti R, Marroni P, Mencoboni M, et al. Individual predictors of increased serum mesothelin in asbestos-exposed workers. *Med Oncol* 2013;30:422. doi: 10.1007/s12032-012-0422-6.

15. Garritano S, De Santi C, Silvestri R, et al. A common polymorphism within MSLN affects miR-611 binding site and soluble mesothelin levels in healthy people. *J Thorac Oncol* 2014;9:1662-1668. doi: 10.1097/JTO.0000000000000322.

16. Cristaudo A, Foddis R, Bonotti A, et al. Two novel polymorphisms in 5' flanking region of the mesothelin gene are associated with soluble mesothelin-related peptide (SMRP) levels. *Int J Biol Markers* 2011;26:117-123. doi: 10.5301/JBM.2011.8332.

17. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580-585. doi: 10.1038/ng.2653.

18. Ke Y, Reddel RR, Gerwin BI, et al. Establishment of a human in vitro mesothelial cell model system for investigating mechanisms of asbestos-induced mesothelioma. *Am J Pathol* 1989;134:979-991.

19. Cristaudo A, Foddis R, Bonotti A, et al. Polymorphisms in the putative micro-RNA-binding sites of mesothelin gene are associated with serum levels of mesothelin-related protein. *Occup Environ Med* 2010;67:233-236. doi: 10.1136/oem.2009.049205.

20. Saeed HM, Alanazi MS, Alshahrani O, et al. Matrix metalloproteinase-2 C(-1306)T promoter polymorphism and breast cancer risk in the Saudi population. *Acta Biochim Pol* 2013;60:405-409.
21. Wu H, Zhang K, Gong P, et al. A novel functional TagSNP Rs7560488 in the DNMT3A1 promoter is associated with susceptibility to gastric cancer by modulating promoter activity. *PLoS One* 2014;9:e92911. doi: 10.1371/journal.pone.0092911.
22. Cingeetham A, Vuree S, Jiwatani S, et al. Role of the MDM2 promoter polymorphism (-309T>G) in acute myeloid leukemia development. *Asian Pac J Cancer Prev* 2015;16:2707-2012.
23. Domingues-Montanari S, Subirana I, Tomás M, et al. Association between ESR2 genetic variants and risk of myocardial infarction. *Clin Chem* 2008;54:1183-1189. doi: 10.1373/clinchem.2007.102400.
24. Singh K, Agrawal NK, Gupta SK, et al. 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* 2013;12:199-204. doi: 10.1177/1534734613493289.
25.
25. Park JH, Lee J, Kim CH, et al. The polymorphism (-600 C>A) of CpG methylation site at the promoter region of CYP17A1 and its association of male infertility and testosterone levels. *Gene* 2014;534:107-12. doi: 10.1016/j.gene.2013.09.088.
26. de Toro-Martín J, Guénard F, Tchernof A, et al. A CpG-SNP Located within the ARPC3 Gene Promoter Is Associated with Hypertriglyceridemia in Severely Obese Patients. *Ann Nutr Metab* 2016;68:203-12. doi: 10.1159/000445358.
27. Tan K, Kajino K, Momose S, et al. Mesothelin (MSLN) promoter is hypomethylated in malignant mesothelioma, but its expression is not associated with methylation status of the promoter. *Hum Pathol* 2010;41:1330-8. doi: 10.1016/j.humpath.2010.03.002.

28. Nelson HH, Almquist LM, LaRocca JL, et al. The relationship between tumor MSLN methylation and serum mesothelin (SMRP) in mesothelioma. *Epigenetics* 2011;6:1029-34. doi: 10.4161/epi.6.8.16074.

29. Cramer SD, Sun J, Zheng SL, et al. Association of prostate-specific antigen promoter genotype with clinical and histopathologic features of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17:2451-2457. doi: 10.1158/1055-9965.EPI-08-0374.

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. In the figure key, “AA” indicate subjects carrying the common homozygote genotype, “Aa” the heterozygotes, “aa” the variant homozygotes. 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. In each panel the different haplotype constructs of MSLN promoter (left) and dual-luciferase reporter assays results (right) are reported. Left: diagram of four promoter constructs showing four different haplotypes. The number indicates the relative nucleotide position of the four SNPs from the TSS. Right: luciferase activity was presented as the ratio between the firefly/Renilla luciferase values (RLU). RLU measured after pGL3_HAP1 transfection is reported as 100% and used as reference for statistical evaluation. Asterisks show a statistical significance ($p<0.05$) in the Dunnett test for pairwise differences within the analysis of variance (ANOVA) model between pGL3_HAP2/3/4 and the reference pGL3_HAP1 in panels A and B. Asterisks show a statistical significance ($p<0.05$) in the Sidak test for pairwise differences within the multivariate ANOVA model between pGL3_HAP2/3/4 and the reference pGL3_HAP1 in panel C. The columns represent mean values, the bars show the standard error of the mean (SEM). MPM, malignant pleural mesothelioma; SNPs, single nucleotide polymorphisms; TSS, transcriptional start site.

Figure 1

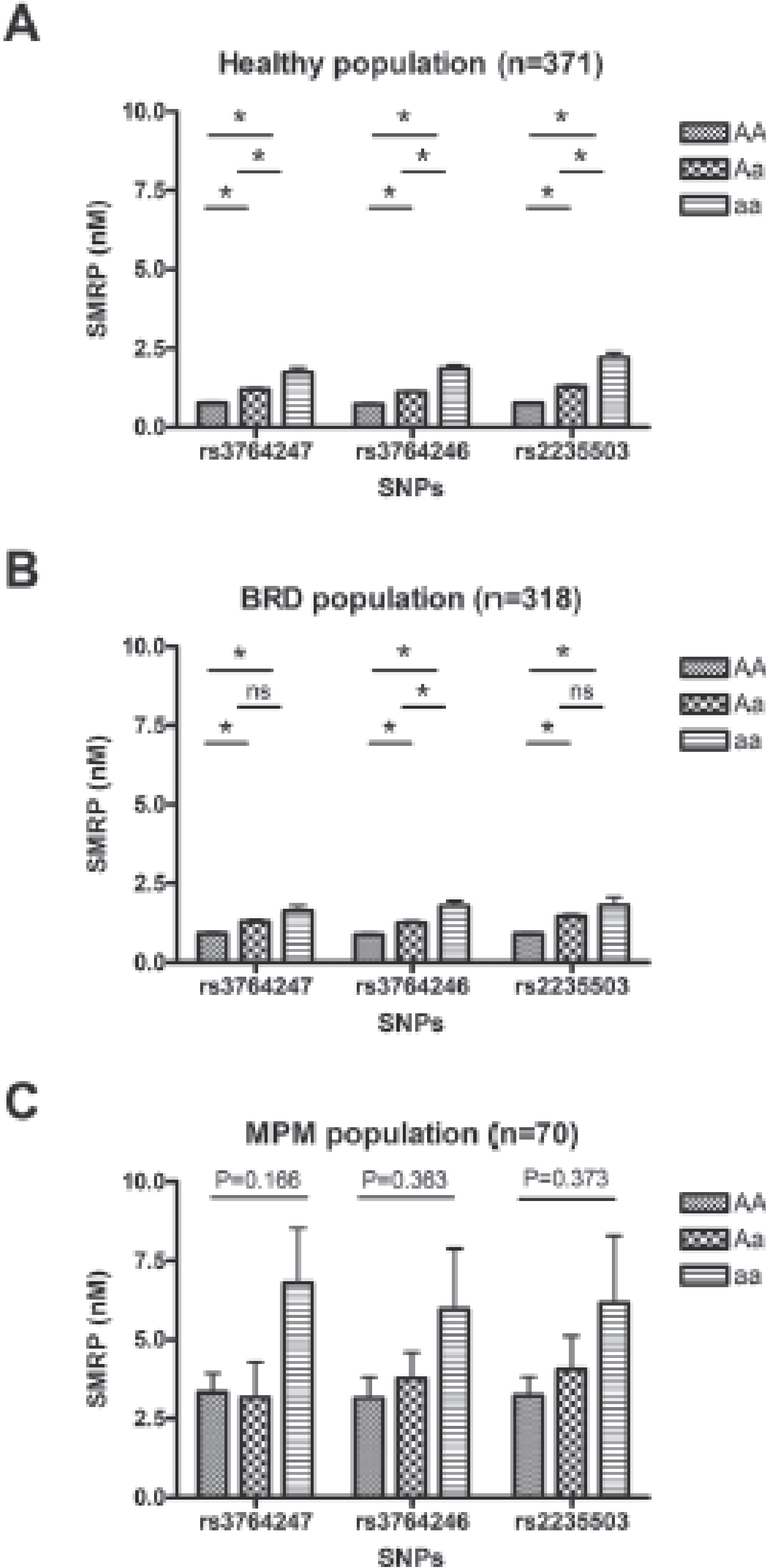

Figure 2