



# Variation rs2235503 C > A Within the Promoter of *MSLN* Affects Transcriptional Rate of Mesothelin and Plasmatic Levels of the Soluble Mesothelin-Related Peptide

Roberto Silvestri<sup>1</sup>, Perla Pucci<sup>1,2</sup>, Chiara De Santi<sup>1,3</sup>, Irene Dell'Anno<sup>1</sup>, Simona Miglietta<sup>1,4</sup>, Alda Corrado<sup>1,5</sup>, Vanessa Nicolì<sup>1,6</sup>, Daniela Marolda<sup>1</sup>, Monica Cipollini<sup>1</sup>, Enrica Pellegrino<sup>1</sup>, Monica Evangelista<sup>7</sup>, Alessandra Bonotti<sup>8</sup>, Rudy Foddìs<sup>6</sup>, Alfonso Cristaudo<sup>6</sup>, Stefano Landi<sup>1\*</sup> and Federica Gemignani<sup>1</sup>

## OPEN ACCESS

### Edited by:

Changning Liu,  
Xishuangbanna Tropical Botanical  
Garden (CAS), China

### Reviewed by:

Jingjing Liu,  
St. Jude Children's Research  
Hospital, United States  
Xiangqian Guo,  
Henan University, China

### \*Correspondence:

Stefano Landi  
slandi@biologia.unipi.it

### Specialty section:

This article was submitted to  
Genomic Medicine,  
a section of the journal  
Frontiers in Genetics

**Received:** 05 May 2020

**Accepted:** 31 July 2020

**Published:** 18 August 2020

### Citation:

Silvestri R, Pucci P, De Santi C, Dell'Anno I, Miglietta S, Corrado A, Nicolì V, Marolda D, Cipollini M, Pellegrino E, Evangelista M, Bonotti A, Foddìs R, Cristaudo A, Landi S and Gemignani F (2020) Variation rs2235503 C > A Within the Promoter of *MSLN* Affects Transcriptional Rate of Mesothelin and Plasmatic Levels of the Soluble Mesothelin-Related Peptide. *Front. Genet.* 11:975. doi: 10.3389/fgene.2020.00975

<sup>1</sup> Department of Biology, University of Pisa, Pisa, Italy, <sup>2</sup> Division of Cellular and Molecular Pathology, Department of Pathology, University of Cambridge, Cambridge, United Kingdom, <sup>3</sup> Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland, <sup>4</sup> San Raffaele Telethon Institute for Gene Therapy (SR-Tiget), San Raffaele Scientific Institute (IRCCS), Milan, Italy, <sup>5</sup> Department of Bioscience, University of Milan, Milan, Italy, <sup>6</sup> Department of Translational Research and of New Surgical and Medical Technologies, University of Pisa, Pisa, Italy, <sup>7</sup> Institute of Clinical Physiology (IFC), CNR, Pisa, Italy, <sup>8</sup> Preventive and Occupational Medicine, University Hospital of Pisa, Pisa, Italy

Soluble mesothelin-related peptide (SMRP) is a promising biomarker for malignant pleural mesothelioma (MPM), but several confounding factors can reduce SMRP-based test's accuracy. The identification of these confounders could improve the diagnostic performance of SMRP. In this study, we evaluated the sequence of 1,000 base pairs encompassing the minimal promoter region of the *MSLN* gene to identify expression quantitative trait loci (eQTL) that can affect SMRP. We assessed the association between four *MSLN* promoter variants and SMRP levels in a cohort of 72 MPM and 677 non-MPM subjects, and we carried out *in vitro* assays to investigate their functional role. Our results show that rs2235503 is an eQTL for *MSLN* associated with increased levels of SMRP in non-MPM subjects. Furthermore, we show that this polymorphic site affects the accuracy of SMRP, highlighting the importance of evaluating the individual's genetic background and giving novel insights to refine SMRP specificity as a diagnostic biomarker.

**Keywords:** soluble mesothelin-related peptide, malignant pleural mesothelioma, diagnostic biomarkers, single nucleotide polymorphisms, fluorescent reporter assay, eQTL

## INTRODUCTION

Malignant pleural mesothelioma (MPM) is a highly aggressive and rare cancer of the pleura triggered by exposure to asbestos and associated with a poor prognosis (Bianchi et al., 2017). The long latency between the exposure to asbestos and the onset of the disease, the unspecific symptoms at the presentation, and the lack of accurate and non-invasive diagnostic tools make the diagnosis quite challenging (Rusch et al., 2012; Bianco et al., 2018). Thus, MPM is often diagnosed at advanced stages, thereby further reducing the limited therapeutic options available

and resulting in poor prognostic outcomes. The identification of accurate diagnostic biomarkers, or the optimization of those identified so far, could help to overcome this problem (Sun et al., 2017). To date, one of the most promising biomarkers for the diagnosis of MPM is “soluble mesothelin-related peptide” (SMRP), which is released in the extracellular environment following the proteolytic cleavage of mesothelin (a membrane protein encoded by the *MSLN* gene) (Sapede et al., 2008). The plasmatic concentration of SMRP is easily measurable from blood samples, and it is usually higher in MPM than in non-MPM subjects (either healthy or affected by other respiratory conditions), allowing fair discrimination between these two groups (Gao et al., 2019; Gillezeau et al., 2019). Despite these promising features, several confounding factors can reduce the accuracy of SMRP as a diagnostic biomarker, thereby preventing its employment in the clinical practice. While many studies extensively documented the impact of age, body mass index, glomerular filtration rate (Casjens et al., 2017), and tumor histology on SMRP (Scherpereel et al., 2006), none of them explored the role of the genetic background. However, we believe that the genetic background could play a prominent role in this context. An increasing number of studies, by showing the importance of genetic variations in altering the concentration and the accuracy of several biomarkers, strongly support this idea (Cramer et al., 2003; Gudmundsson et al., 2010; Enroth et al., 2014; Ruggiero et al., 2015; Wang et al., 2018). Furthermore, we showed for the first time that single nucleotide polymorphisms (SNPs) located within the 3'-UTR of the *MSLN* gene could affect the SMRP levels through a miRNA-mediated mechanism (Garritano et al., 2014). Recently, we reported similar results for other variants lying within the *MSLN* promoter (De Santi et al., 2017). Based on these observations, we hypothesized that expression quantitative trait loci (eQTL) for *MSLN* could affect the plasmatic concentration and the accuracy of SMRP. Thus, we assessed the association between SNPs located within the promoter region of *MSLN* and the plasmatic levels of SMRP in a cohort of 72 MPM and 677 non-MPM subjects. We carried out *in vitro* assays to investigate the functional role of these SNPs, and highlighted the relationship between the transcriptional rate of *MSLN* gene and serum concentrations of SMRP. Our results show that evaluating the individual's genotype to establish personalized cutoff values can increase the diagnostic accuracy of SMRP, thereby helping to promote its translation into clinical practice.

## MATERIALS AND METHODS

### Selection of the SNPs

Following the hypothesis that the transcriptional rate of the *MSLN* gene could affect the plasmatic concentration of SMRP, we extracted a window of 1,000 base pairs upstream of the transcription starting site (TSS), encompassing the minimal promoter region, using the UCSC genome browser<sup>1</sup>. Within this sequence, we selected those SNPs showing a minor allele

frequency (MAF) >0.05 and a significant association with *MSLN* mRNA levels in lung tissues (according to the GTEx Portal V6p<sup>2</sup>). This selection led to the identification of four SNPs worth of further investigations: rs3764247 A > C, rs3764246 A > G, rs2235503 C > A and rs2235504 A > G (**Supplementary Table 1**). According to 1,000 Genomes<sup>3</sup>, their combination elicits four haplotypes (**Table 1**), accounting for almost the totality (98.5%) of the genetic variability of the promoter region within Caucasians, and Tuscans.

### Genotyping, Haplotype Reconstruction, SMRP Measurements and Association Study

Detailed information about the studied population, DNA extraction, genotyping, and measurement of SMRP levels can be found elsewhere (De Santi et al., 2017). Briefly, we analyzed a total of 677 non-MPM and 72 MPM volunteers. Among the 677 non-MPM, 371 were healthy people, while 306 were patients affected by benign respiratory diseases (BRD). All the subjects were recruited at the University Hospital of Pisa as part of an occupational surveillance program on workers previously exposed to asbestos, as described by Garritano et al. (2014). Blood and serum samples were obtained by venipuncture and store at -80°C. To measure the plasmatic concentration of the SMRP, we employed the Mesomark enzyme-linked immunosorbent assay (Fujirebio Diagnostics, Japan) following the manufacturer's instructions. We carried out DNA extraction and genotyping using the EuroGOLD Blood DNA Mini Kit (EuroClone, Pero, Italy) and the KASPar PCR SNP genotyping system (LGC Genomics Ltd., Teiddington, Middlesex, United Kingdom), respectively. For this study, we reconstructed the individual haplotypes and diplotypes using PHASE 2.1.1 (Stephens et al., 2001; Stephens and Scheet, 2005) and stratified our cohort according to health status and diplotypes. To precisely assess the effect of each haplotype, we restricted the association study to carriers of the haplotype #1 (e.g., H1H2, H1H3, and H1H4), using the H1H1 homozygotes as reference. The number of subjects excluded from the analysis was minimal (19 non-MPM and 6 MPM) and did not hamper the statistical power of the study.

### Cell Lines

For the functional study, we employed a non-malignant SV40-immortalized epithelial cell line (MeT-5A) and an epithelioid malignant mesothelioma cell line (Mero-14). MeT-5A were purchased from ATCC (American Type Culture Collection, CRL-9444) and cultured in Medium-199 (Gibco, Life Technologies, Monza, Italy) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, 3 nM epidermal growth factor (EGF), 400 nM hydrocortisone and 870 nM insulin. Mero-14 were kindly donated by Istituto tumori di Genova (National Research Council, Genova, Italy) and cultured in Dulbecco's modified Eagle's medium (DMEM; EuroClone, Pero, Italy) supplemented with 10% FBS and 1% penicillin/streptomycin. Both cell lines were maintained at 37°C and 5% CO<sub>2</sub>.

<sup>1</sup><https://genome.ucsc.edu/>

<sup>2</sup><https://gtexportal.org/>

<sup>3</sup><https://www.internationalgenome.org/1000-genomes-browsers/>

## Construction of Plasmids

In the first part of this work, we aimed to evaluate the role of the four haplotypes herein identified in affecting the activity of the *MSLN* promoter. Thus, we constructed four vectors to employ in the functional study. Each of these vectors harbored one of the haplotypic variants immediately upstream of a green fluorescent protein (GFP) coding sequence. The same vectors also harbored a red fluorescent protein (RFP) coding sequence controlled by the constitutive eukaryotic promoter EF1 (**Figure 1A**). For the cloning procedure, we employed the CloneEZ PCR cloning kit (GenScript, Piscataway, United States). We used the HR220PA-1 vector (System Bioscience, Palo Alto, United States) linearized with *Bst*BI (New England BioLabs, Ipswich, United States) and DNA fragments representing the four haplotypic variants of the *MSLN* promoter. Each of these fragments was obtained by PCR using as a template the genomic DNA of four subjects homozygote for one of the four selected haplotypes (now on abbreviated as H1–H4 when referring to the human genomic DNA, or HAP1–HAP4 when referring to the cloned fragment). The resulting vectors were named HR\_HAP1 to HR\_HAP4. In the second part of this work, we aimed to ascertain the individual role of each of the selected SNPs. To this end, we created three additional vectors, each harboring the uncommon variant of only one of the four SNPs. These vectors were named HR\_246, HR\_503, and HR\_504 and were obtained from HR\_HAP1 by site-directed mutagenesis using the Quick Change Lightning Site-Directed Mutagenesis Kit (Agilent). Consistently with the terminology employed in this second phase of the study, since the HR\_HAP4 differs from the HR\_HAP1 only for the rs3764247, bearing the C-allele instead of the A-allele, we renamed it “HR\_247.” **Figure 1A** shows the details of the changes introduced by the site-directed mutagenesis to generate these additional vectors.

## Fluorescent Reporter Assay

To evaluate the effect of the selected haplotypes and SNPs on the transcriptional activity of the *MSLN* promoter, we employed a

fluorescent reporter assay. To this end,  $2 \times 10^5$  cells/ml were electroporated with 10  $\mu$ g of plasmid DNA, seeded into a six-well plate, and incubated at 37°C and 5% CO<sub>2</sub> for 72 h. After the incubation, we harvested the cells by trypsinization and measured the GFP/RFP fluorescence intensity using the BD FACSJazz System (BD Biosciences, Franklin Lakes, United States). We used the untreated sample to set the RFP intensity threshold for the selection of efficiently transfected cells (**Figure 1B**). Similarly, we used cells transfected with the “empty” HR220PA-1 vector (expressing the RFP but not the GFP reporter gene) to establish the average background signal detected by the FITC/GFP channel (**Figure 1C**). These two steps allowed us to restrict the statistical analysis only to the cells that were efficiently transfected, and that showed a GFP signal above the average background (**Figure 1D**; P3). All the electroporation steps were carried out using the Neon Transfection System (Thermo Fisher Scientific, Monza, Italy), and the following parameters: 1,230 V, 30 ms, 2 pulses for MeT-5A and 1,130 V, 30 ms, 2 pulses for Mero-14.

## Statistical Analyses

For the statistical analyses, we employed GraphPad PRISM 7.0 software. We carried out the analysis of variance (ANOVA) for the association study and the multifactor ANOVA (mANOVA) for the functional studies, both followed by Dunnett’s multiple comparison test. To evaluate the diagnostic performance of SMRP along with the optimal cutoff values, we calculated the receiver operating characteristic (ROC) curves employing the same software.

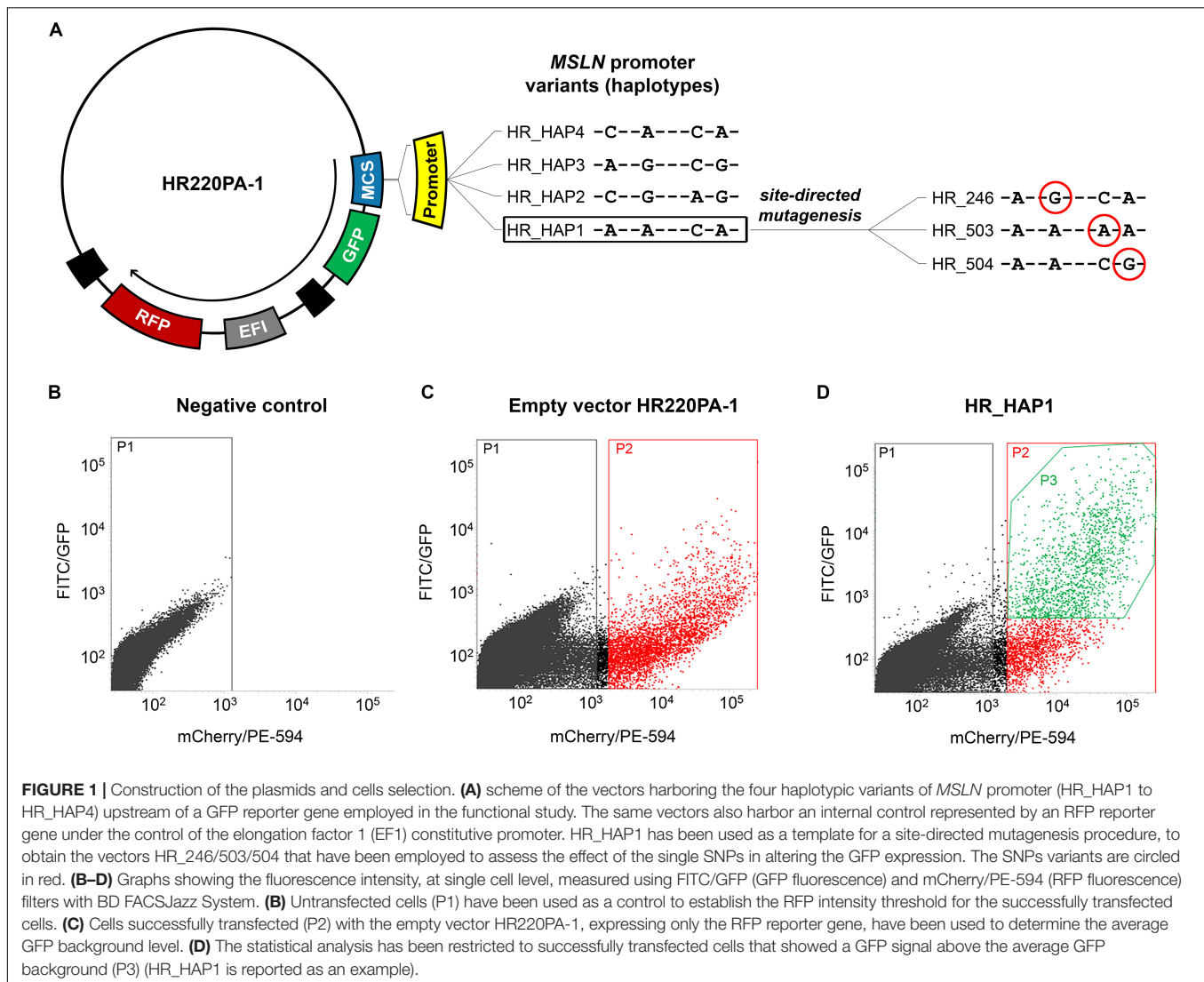
## RESULTS

### Haplotype #2 and Haplotype #3 Are Associated With SMRP Levels *in vivo*

Among the non-MPM volunteers, we found a statistically significant difference between the plasmatic levels of SMRP in heterozygotes H1H2, H1H3, and H1H4 compared with those

**TABLE 1** | Number of individuals stratified according to their health status and genotype at single allele/diplotype level. “MPM” represents patients affected by malignant pleural mesothelioma, while “non-MPM” identifies subjects exposed to asbestos (healthy or affected by benign respiratory diseases). Table also shows the variants (rs) that characterize each haplotype.

Haplotypes distribution and sequences						
Distribution among studied population						
Single Alleles	Non-MPM	MPM	rs3764247	rs3764246	rs2235503	rs2235504
Haplotype 1 (H1)	1047	105	A	A	C	A
Haplotype 2 (H2)	209	24	C	G	A	G
Haplotype 3 (H3)	109	13	A	G	C	G
Haplotype 4 (H4)	47	4	C	A	C	A
Diploypes	Non-MPM	MPM				
H1H1	389	39				
H1H2	152	14				
H1H3	78	11				
H1H4	39	2				



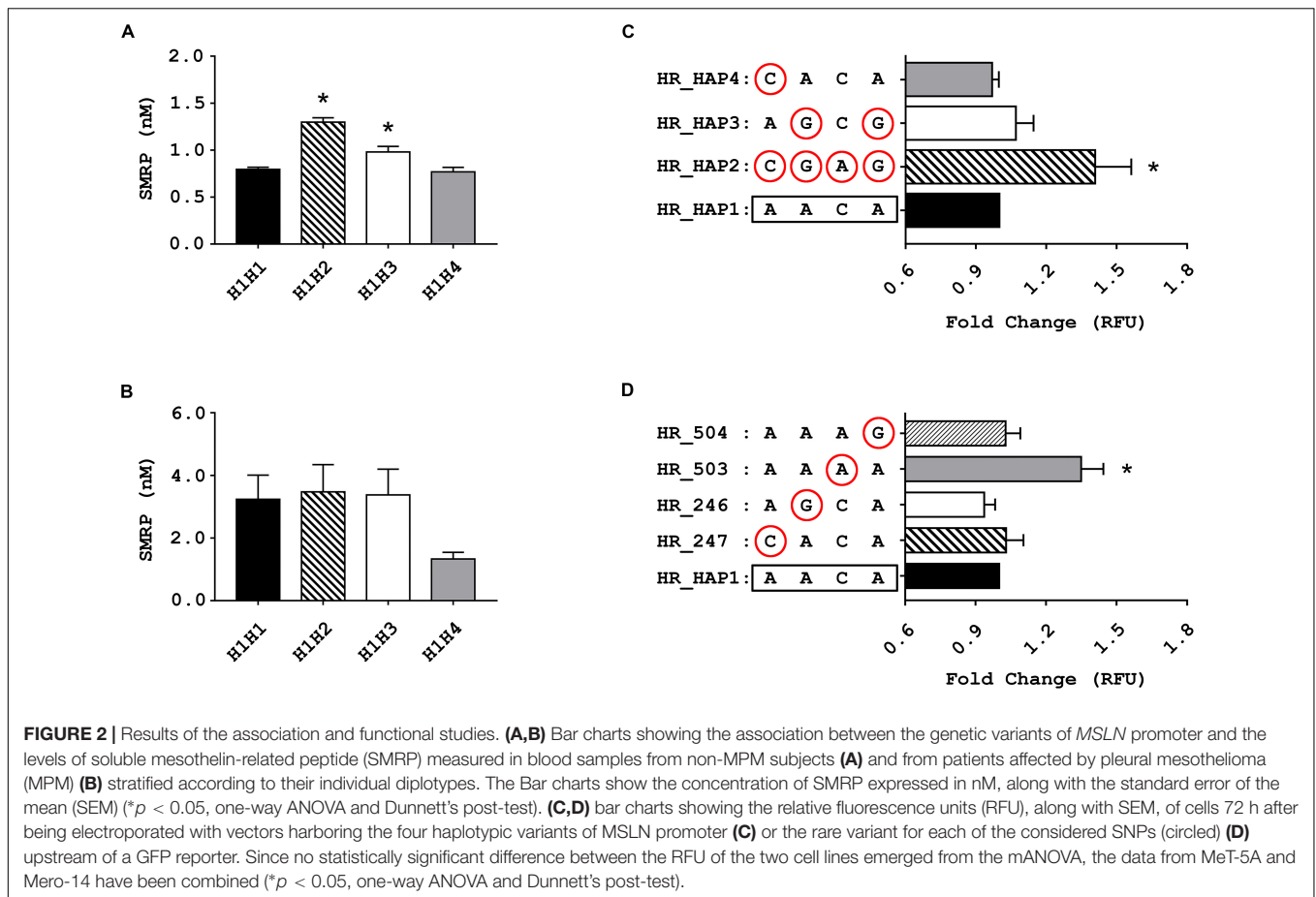
of the reference group H1H1 ( $p$ -value ANOVA  $< 10^{-5}$ ). The Dunnett's post-test showed a statistically significant difference between carriers of haplotype #2 or haplotype #3, but not haplotype #4, and the reference H1H1 ( $p$ -value  $< 10^{-5}$  and 0.0047, respectively) (Figure 2A). On average, carriers of haplotype #2 showed the highest SMRP level (average  $\pm$  standard error:  $1.30 \pm 0.046$  nM) followed by carriers of haplotype #3, #1, and #4 ( $0.98 \pm 0.060$  nM,  $0.80 \pm 0.022$  nM, and  $0.78 \pm 0.047$  nm, respectively). We did not observe any significant difference within the group of MPM patients (Figure 2B).

### ***In vitro* Studies Confirm the Functional Role of Haplotype #2 and Identify rs2235503 C > A as the Most Likely Causative SNP**

To evaluate whether the haplotype #2 and #3 could enhance the activity of the *MSLN* promoter, we carried out *in vitro* experiments using a fluorescence reporter assay. We transfected

MeT-5A and Mero-14 cells with HR vectors carrying the GFP reporter gene under the control of the different haplotypic variants of *MSLN* promoter, named HR\_HAP1, HR\_HAP2, HR\_HAP3, and HR\_HAP4 (Figure 1A). Results showed a statistically significant difference among haplotypes but not between cell lines (mANOVA  $p$ -value = 0.0016 and 0.67). When compared to cells transfected with HR\_HAP1, cells transfected with HR\_HAP2 showed a 1.41-fold increased RFU ( $\pm 0.267$ ;  $p$ -value =  $6 \times 10^{-4}$ ). HR\_HAP3 conferred a slight increased expression, but not statistically significant ( $1.07 \pm 0.131$ ;  $p$ -value = 0.45). HR\_HAP4 showed a promoter activity similar to HR\_HAP1 ( $0.97 \pm 0.05$ ;  $p$ -value = 0.98) (Figure 2C). Since haplotype #2 carries the uncommon variants of the four SNPs, we investigated the functional role of each of them, by employing four more constructs: HR\_247, HR\_246, HR\_503, and HR\_504 (Figure 1A). Each of these vectors bore the uncommon variant of rs3764247, rs3764246, rs2235503, or rs2235504, thus differing from HR\_HAP1 only for one SNP. Again, mANOVA showed a statistically significant difference among genotypes



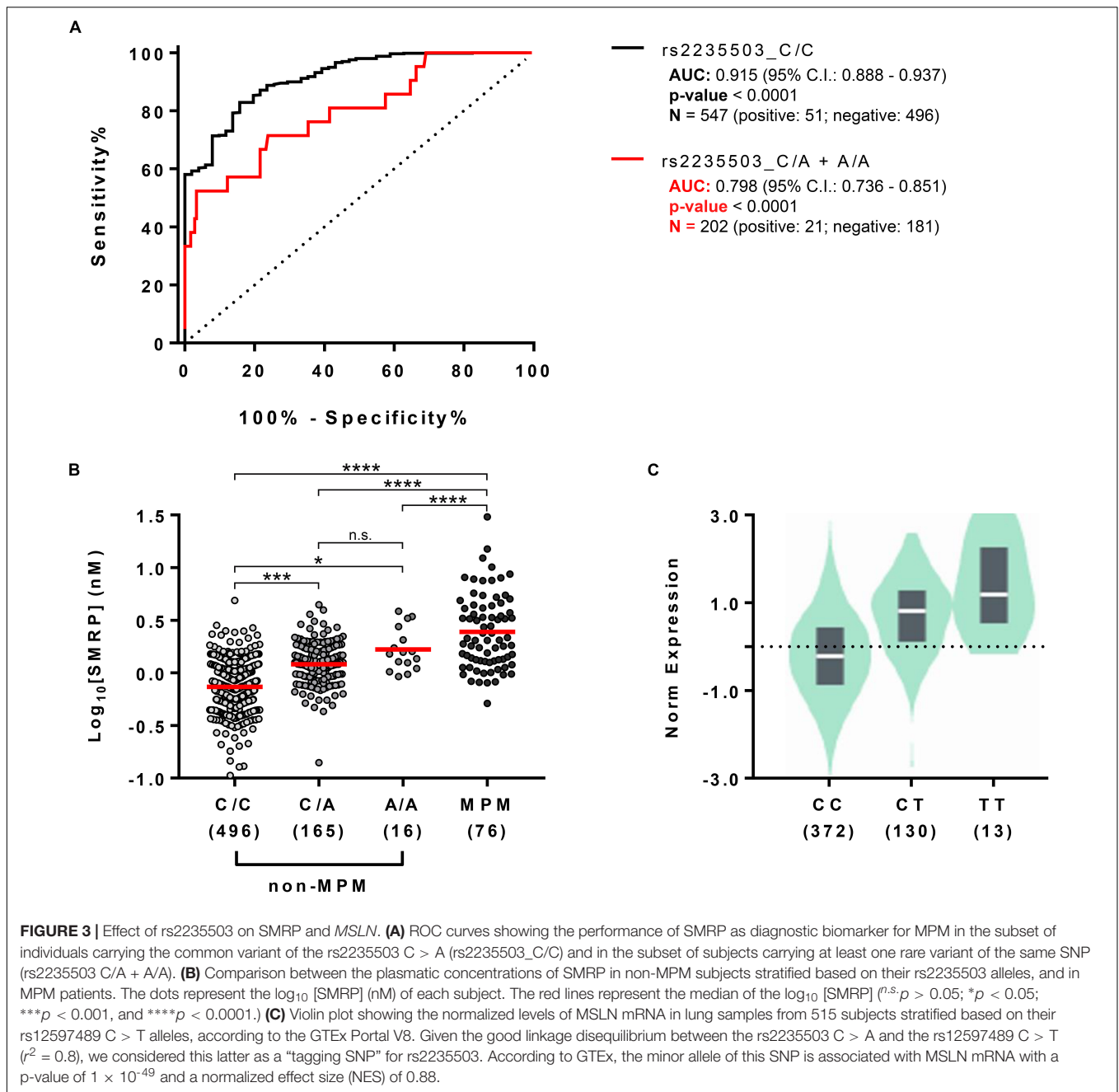


but not between cell lines ( $p$ -value  $< 10^{-5}$  and 0.15). When compared with HR\_HAP1, only cells transfected with HR\_503 showed a significant RFU increase (fold change  $1.35 \pm 0.164$ ;  $p$ -value  $< 10^{-5}$ ) (Figure 2D).

### Rs2235503 C > A Significantly Affects the Performance of SMRP as Diagnostic Biomarker for MPM

Since the results indicated that the uncommon variant A-rs2235503 caused an increase in *MSLN* expression and SMRP levels in non-MPM subjects, we sought to determine whether this effect could significantly affect the accuracy of SMRP as a diagnostic biomarker for MPM. To verify this aspect, we stratified our cohort in two subgroups, the former containing the carriers of A-rs2235503 (rs2235503\_C/A + A/A) and the latter containing all the other subjects (rs2235503\_C/C). We then used the SMRP levels to calculate the “ROC” curve for each of the two subgroups. Results suggested a strong influence of the genotype on the performance of SMRP. In fact, the ROC curve for the C/A + A/A group resulted in an area under the curve (AUC) of only  $0.798 \pm 0.055$  compared with an AUC of  $0.915 \pm 0.018$  for the C/C group (Figure 3A). Interestingly, the AUC for the C/C group was also higher than that for the overall population ( $0.877 \pm 0.019$ ). Moreover, the

Youden index pointed at an optimal cut-off value of 1.118 (sensitivity = 84.31, specificity = 82.86) for the C/C group, and 3.092 (sensitivity = 52.58, specificity = 96.69) for the C/A + A/A group. The optimal cut-off value for the overall population was 1.28, with a sensitivity of 77.78 and a specificity of 79.91. Notably, using a cut-off of 1.28 nM, the specificity for the C/C group raised to 88.71, but the sensitivity dropped to 76.47. Similarly, the specificity for the C/A + A/A group raised to 55.80, but the sensitivity dropped to 80.95. **Supplementary Tables 2–4** report the complete list of the possible cut-off values along with the associated sensitivity and specificity for each group. These results were not surprising as the difference in the average SMRP concentration between non-MPM and MPM subjects, although still significant ( $p$ -value  $< 0.0001$ ), dropped from  $2.734 \pm 0.174$  nM for the C/C group to only  $1.706 \pm 0.388$  nM for the A/A group (Figure 3B). Intriguingly, data from GTEx Portal (V8) showed a similar trend for the tagging SNP rs12597489 C > T, in linkage disequilibrium (LD) with rs2235503 ( $r^2 = 0.8$  according to HaploReg v4.1), suggesting a strong relationship between the mRNA levels of *MSLN* and the serum concentration of SMRP (Figure 3C). To explore the potential relationship between the rs2235503 and the prognosis of MPM patients, we carried out a survival analysis. The results showed a lack of association between this SNP and the overall survival ( $p$ -value = 0.66). The multivariate analysis performed



to evaluate the prognostic significance of rs2235503, SMRP, and histology led to similar results, with none of these factors being significantly associated with prognosis (**Supplementary Figures 1A,B**).

## DISCUSSION

Soluble mesothelin-related peptide is one of the most promising biomarkers for MPM, but its sensitivity and specificity are suboptimal for use in the clinical practice (Scherpereel et al., 2006; Casjens et al., 2017). Our previous data showed that SNPs

within regulatory regions of the *MSLN* gene were associated with SMRP levels, raising the question of whether the evaluation of these variants could improve the performance of SMRP in the diagnosis of MPM (Garritano et al., 2014; De Santi et al., 2017). Herein, we explored the possibility of exploiting eQTL for *MSLN* to increase the diagnostic accuracy of SMRP. Firstly, we carried out an association study in a cohort of 677 non-MPM subjects and 72 MPM patients. We found that two haplotypes (#2 and #3) of the minimal promoter region of the *MSLN* gene were associated with increased serum levels of SMRP in non-MPM subjects. In agreement with our previous studies, we did not observe any association in the cohort of MPM

patients (Garritano et al., 2014; De Santi et al., 2017). While we cannot rule out that the lack of association observed in MPM patients could be ascribed to a limited statistical power ( $n = 72$ ), Goricar et al. (2020) recently reported similar results, showing that the minor allele of the rs1057147 was associated with increased SMRP levels in non-MPM ( $n = 782$ ) but not in MPM subjects ( $n = 154$ ). Together, these observations seem to corroborate the hypothesis that genotype has only a limited effect on SMRP levels in the malignant context, probably due to the major role played by other cancer-related factors in altering SMRP concentration. However, more studies on other cohorts of MPM patients are warranted for better exploring this effect among affected people (Pass et al., 2008). Nonetheless, the *in vitro* assays confirmed the functional role of haplotype #2, able to increase the expression of a reporter gene in a chimerized vector of about 42%. Moreover, cells transfected with a vector harboring the A-rs2235503 showed increased promoter activity of a similar extent, compared to C-rs2235503. This result straightened the role of haplotype #2 and pinpointed rs2235503 as the SNP responsible for the enhancement of *MSLN* gene expression. Since SMRP derives from the proteolytic cleavage of mesothelin (Sapede et al., 2008), it is conceivable that its plasmatic levels can be influenced by factors such as the eQTL for *MSLN*. Our work supports the rationale that the polymorphism rs2235503 is among these factors. Unfortunately, we could not perform any measurement of mRNA *MSLN* expression on the cohort of volunteers recruited in this study, however, indirectly, also the data from 525 lung samples within GTEx biobank confirmed a relationship between the genotype and the extent of *MSLN* gene transcription. Conversely, our *in vitro* studies ruled out the possibility of a functional role for the SNPs that characterize the haplotype #3 (i.e., rs3764246 and rs2235504). Thus, likely, the association observed *in vivo* should be ascribed to other SNPs in LD with haplotype #3 but residing outside the 1,000-bps promoter region herein considered. Since the results from other groups suggest the possibility of a prognostic significance of SMRP-affecting SNPs (Dipalma et al., 2011; Goricar et al., 2020), we tested whether genotype and SMRP levels could be predictive of a worse prognosis in our cohort. Our observations seem to rule out this hypothesis. Given the flaw association between SNP and SMRP among patients, and the lack of association between SMRP and overall survival, it is not surprising that there is a lack of significant relationship between SNP and patients' survival in our cohort. However, given the limited number of MPM patients recruited in the present study, this result should be carefully evaluated and more work involving additional cohorts is needed to elucidate the prognostic role of the rs2235503. Then, we assessed whether the performance of SMRP could be improved when taking into account the subjects' genetic background, in agreement with previous findings for other biomarkers (Cramer et al., 2003; Gudmundsson et al., 2010; Enroth et al., 2014; Ruggiero et al., 2015; Wang et al., 2018). Notably, we found that the AUC for individuals carrying the rs2235503-C/C genotype was higher (0.915) than that of carriers of the A-allele (0.798) and higher as compared to the whole population (0.877). Moreover, the Youden index pointed at significantly different cut-off values according to the rs2235503

genotypes, suggesting that the stratification of subjects based on this SNP could improve the accuracy of SMRP. In conclusion, our work shows that rs2235503 affects *MSLN* gene transcription and represents an eQTL for SMRP levels. Moreover, we highlighted for the first time that the rs2235503 can affect the diagnostic performance of SMRP, reinforcing the importance of considering the individual genetic background to improve the accuracy of cancer biomarkers (Cramer et al., 2003; Gudmundsson et al., 2010; Enroth et al., 2014; Wang et al., 2018). Therefore, it is conceivable that the characterization of further eQTL could help to translate SMRP into the clinical practice and improve the efficiency of this biomarker.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GTEx Portal (<https://gtexportal.org/home/index.html>) and dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee of the University Hospital of Pisa. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SL and FG conceived the idea at the basis of the study, supervised the development of the work, and revised the manuscript. AB, RF, and AC from the Occupational Medicine of Cisanello Hospital in Pisa, contributed with essential biological samples and the related clinical features. CD carried out DNA extraction and genotyping. RS and PP obtained the plasmids via cloning and mutagenesis and performed the functional studies with the contribution of SM, AC, VN, ID, DM, EP, and MC. ME carried out the cytofluorimetric analyses. RS wrote the manuscript. PP, CD, VN, ID, DM, SL, and FG contributed to the revision of the manuscript. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

A preprint version of this manuscript is available on Authorea (Silvestri et al., 2020).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00975/full#supplementary-material>

## REFERENCES

- Bianchi, C., Bianchi, T., Pass, H. I., Wali, A., Tang, N., Ivanova, A., et al. (2017). Global mesothelioma epidemic: trend and features. *J. Thorac. Oncol.* 9, 82–88.
- Bianco, A., Valente, T., De Rimini, M. L., Sica, G., and Fiorelli, A. (2018). Clinical diagnosis of malignant pleural mesothelioma. *J. Thorac. Dis.* 10(Suppl. 2), S253–S261.
- Casjens, S., Weber, D. G., Johnen, G., Raiko, I., Taeger, D., Meinig, C., et al. (2017). Assessment of potential predictors of calretinin and mesothelin to improve the diagnostic performance to detect malignant mesothelioma: results from a population-based cohort study. *BMJ Open* 7:e017104. doi: 10.1136/bmjopen-2017-017104
- Cramer, S. D., Chang, B.-L., Rao, A., Hawkins, G. A., Zheng, S. L., Wade, W. N., et al. (2003). Association between genetic polymorphisms in the prostate-specific antigen gene promoter and serum prostate-specific antigen levels. *J. Natl. Cancer Inst.* 95, 1044–1053. doi: 10.1093/jnci/95.14.1044
- De Santi, C., Pucci, P., Bonotti, A., Melaiu, O., Cipollini, M., Silvestri, R., et al. (2017). Mesothelin promoter variants are associated with increased soluble mesothelin-related peptide levels in asbestos-exposed individuals. *Occup. Environ. Med.* 74, 456–463.
- Dipalma, N., Luisi, V., Di Serio, F., Fontana, A., Maggiolini, P., Licchelli, B., et al. (2011). Biomarkers in malignant mesothelioma: diagnostic and prognostic role of soluble mesothelin-related peptide. *Int. J. Biol. Markers* 26, 160–165. doi: 10.5301/jbm.2011.8614
- Enroth, S., Johansson, A., Enroth, S. B., and Gyllensten, U. (2014). Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat. Commun.* 5:4684.
- Gao, R., Wang, F., Wang, Z., Wu, Y., Xu, L., Qin, Y., et al. (2019). Diagnostic value of soluble mesothelin-related peptides in pleural effusion for malignant pleural mesothelioma: an updated meta-analysis. *Medicine* 98:e14979. doi: 10.1097/md.00000000000014979
- Garritano, S., De Santi, C., Silvestri, R., Melaiu, O., Cipollini, M., Barone, E., et al. (2014). A common polymorphism within MSLN affects miR-611 binding site and soluble mesothelin levels in healthy people. *J. Thorac. Oncol.* 9, 1662–1668. doi: 10.1097/jto.0000000000000322
- Gillezeau, C., van Gerwen, M., Ramos, J., Liu, B., Flores, R., and Taioli, E. (2019). Biomarkers for malignant pleural mesothelioma: a meta-analysis. *Carcinogenesis* 40, 1320–1331. doi: 10.1093/carcin/bg z103
- Goricar, K., Kovac, V., Dodic-Fikfak, M., Dolzan, V., and Franko, A. (2020). Evaluation of soluble mesothelin-related peptides and MSLN genetic variability in asbestos-related diseases. *Radiol. Oncol.* 54, 86–95. doi: 10.2478/raon-2020-0011
- Gudmundsson, J., Besenbacher, S., Sulem, P., Gudbjartsson, D. F., Olafsson, I., Arinbjarnarson, S., et al. (2010). Genetic correction of PSA values using sequence variants associated with PSA levels. *Sci. Transl. Med.* 2:62ra92.
- Pass, H. I., Wali, A., Tang, N., Ivanova, A., Ivanov, S., Harbut, M., et al. (2008). Soluble mesothelin-related peptide level elevation in mesothelioma serum and pleural effusions. *Ann. Thorac. Surg.* 85, 265–272. doi: 10.1016/j.athoracsur.2007.07.042
- Ruggiero, D., Nappo, S., Nutile, T., Sorice, R., Talotta, F., Giorgio, E., et al. (2015). Genetic variants modulating CRIPTO serum levels identified by genome-wide association study in Cilento isolates. *PLoS Genet.* 11:e1004976. doi: 10.1371/journal.pgen.1004976
- Rusch, V. W., Giroux, D., Kennedy, C., Ruffini, E., Cangir, A. K., Rice, D., et al. (2012). Initial analysis of the international association for the study of lung cancer mesothelioma database. *J. Thorac. Oncol.* 7, 1631–1639. doi: 10.1097/jto.0b013e31826915f1
- Sapede, C., Gauvrit, A., Barbieux, I., Padieu, M., Cellerin, L., Sagan, C., et al. (2008). Aberrant splicing and protease involvement in mesothelin release from epithelioid mesothelioma cells. *Cancer Sci.* 99, 590–594. doi: 10.1111/j.1349-7006.2007.00715.x
- Scherpereel, A., Grigoriu, B., Conti, M., Gey, T., Gregoire, M., Copin, M.-C., et al. (2006). Soluble mesothelin-related peptides in the diagnosis of malignant pleural mesothelioma. *Am. J. Respir. Crit. Care Med.* 173, 1155–1160. doi: 10.1164/rccm.200511-1789oc
- Silvestri, R., Pucci, P., Santi, C., De Anno, I. D., Miglietta, S., Corrado, A., et al. (2020). Variation rs2235503 C>A Within the Promoter of MSLN Affects Transcriptional rate of Mesothelin and Plasmatic Levels of the Soluble Mesothelin-Related Peptide. Hoboken, NJ: Authorea, Inc.
- Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76, 449–462. doi: 10.1086/428594
- Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989. doi: 10.1086/319501
- Sun, H. H., Vaynblat, A., and Pass, H. I. (2017). Diagnosis and prognosis-review of biomarkers for mesothelioma. *Ann. Transl. Med.* 5:244. doi: 10.21037/atm.2017.06.60
- Wang, K., Bai, Y., Chen, S., Huang, J., Yuan, J., Chen, W., et al. (2018). Genetic correction improves prediction efficiency of serum tumor biomarkers on digestive cancer risk in the elderly Chinese cohort study. *Oncotarget* 9, 7389–7397. doi: 10.18632/oncotarget.23205

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Silvestri, Pucci, De Santi, Dell'Anno, Miglietta, Corrado, Nicoli, Marolda, Cipollini, Pellegrino, Evangelista, Bonotti, Foddis, Cristaudo, Landi and Gemignani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.