
Copyright: © 2013 Matullo et al.

This is the final published version of the article accepted for publication in PloS One following peer review. The definitive published version (see citation above) is located on the article abstract page of the publisher, Public Library of Science.

This version was made available in the UWA Research Repository on 16th January 2014, in compliance with the publisher’s policies on archiving in institutional repositories.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Genetic Variants Associated with Increased Risk of Malignant Pleural Mesothelioma: A Genome-Wide Association Study

Giuseppe Matullo1,2,*, Simonetta Guarrera1, Marta Betti3, Giovanni Fiorito1, Daniela Ferrante4, Floriana Voglino1, Gemma Cadby5,6,7, Cornelia Di Gaetano1,2, Fabio Rosa1, Alessia Russo1,2, Ari Hirvonen8, Elisabetta Casalone3, Sara Tunesi4, Marina Padoan4, Mara Giordano9, Anna Aspesi9, Caterina Casadiso10, Francesco Ardissone11, Enrico Ruffini12, Pier Giacomo Betta13, Roberto Libener13, Roberto Guaschino14, Ezio Piccolini15, Monica Neri16, Arthur W. B. Musk17,18, Nicholas H. de Klerk19, Jennie Hui18,20, John Beilby18,20, Alan L. James17,18, Jenette Creaney17,18, Bruce W. Robinson17,18, Sutapa Mukherjee21,22, Lyle J. Palmer5,6, Dario Mirabelli23,24, Donatella Ugolini25, Stefano Bonassi16, Corrado Magnani4,24, Irma Dianzani3,24

1 Human Genetics Foundation, HUGeF, Turin, Italy, 2 Department of Medical Sciences, University of Turin, Turin, Italy, 3 Laboratory of Genetic Pathology, Department of Health Sciences, University of Piemonte Orientale, Novara, Italy, 4 CPO-Piemonte and Unit of Medical Statistics and Epidemiology, Department Translational Medicine, University of Piemonte Orientale, Novara, Italy, 5 Genetic Epidemiology and Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario, Canada, 6 Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada, 7 Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Nedlands, Western Australia, Australia, 8 Centre of Expertise for Health and Work Ability, Finnish Institute of Occupational Health, Helsinki, Finland, 9 Laboratory of Genetics, Department Health Sciences, University of Piemonte Orientale, Novara, Italy, 10 Thoracic Surgery Unit, University of Piemonte Orientale, Novara, Italy, 11 Chest Surgery, Department of Clinical and Biological Sciences, University of Turin, Orbassano, Italy, 12 Thoracic Surgery Unit, University of Turin, Turin, Italy, 13 Pathology Unit, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, 14 Transfusion Centre, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, 15 Pneumology Unit, Santo Spirito Hospital, Casale Monferrato, Italy, 16 Unit of Clinical and Molecular Epidemiology ERCSS San Raffaele Pisana, Rome, Italy, 17 Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia, 18 National Centre for Asthma Related Disease, School of Medicine and Pharmacology, University of Western Australia, Nedlands, Western Australia, 19 Centre for Child Health Research, The University of Western Australia, Nedlands, Western Australia, Australia, 20 PathWest Laboratory Medicine WA, Nedlands, Western Australia, Australia, 21 Department of Medicine, University of Toronto, Toronto, Ontario, Canada, 22 Women’s College Research Institute and Women’s College Hospital, Toronto, Ontario, Canada, 23 Unit of Cancer Epidemiology, CPO-Piemonte and University of Turin, Turin, Italy, 24 Interdepartmental Center for Studies on Asbestos and other Toxic Particulates “G. Scarselli”, University of Turin, Turin, Italy, 25 Department of Internal Medicine, University of Genoa and IRCCS AOI San Martino-IST-Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy

Abstract

Asbestos exposure is the main risk factor for malignant pleural mesothelioma (MPM), a rare aggressive tumor. Nevertheless, only 5–17% of those exposed to asbestos develop MPM, suggesting the involvement of other environmental and genetic risk factors. To identify the genetic risk factors that may contribute to the development of MPM, we conducted a genome-wide association study (GWAS; 370,000 genotyped SNPs, 5 million imputed SNPs) in Italy, among 407 MPM cases and 1269 controls with a complete history of asbestos exposure. A replication study was also undertaken and included 428 MPM cases and 1269 controls from Australia. Although no single marker reached the genome-wide significance threshold, several associations were supported by haplotype-, chromosomal region-, gene- and gene-ontology process-based analyses. Most of these SNPs were located in regions reported to harbor aberrant alterations in mesothelioma (SLC7A14, THRB, CEBP350, ADAMTS2, ETV1, PVT1 and MMP14 genes), causing at most a 2–3-fold increase in MPM risk. The Australian replication study showed significant associations in five of these chromosomal regions (3q26.2, 4q32.1, 7p22.2, 14q11.2, 15q14). Multivariate analysis suggested an independent contribution of 10 genetic variants, with an Area Under the ROC Curve (AUC) of 0.76 when only exposure and covariates were included in the model, and of 0.86 when the genetic component was also included, with a substantial increase of asbestos exposure risk estimation (odds ratio, OR: 45.28, 95% confidence interval, CI: 21.52–95.28). These results showed that genetic risk factors may play an additional role in the development of MPM, and that these should be taken into account to better estimate individual MPM risk in individuals who have been exposed to asbestos.


Editor: Xiao-Ping Miao, MOE Key Laboratory of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China

Received January 7, 2013; Accepted March 6, 2013; Published April 23, 2013

Copyright: © 2013 Matullo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Regione Piemonte Ricerca Sanitaria Finalizzata 2007, 2008, 2009 (to I.D.), Fondazione Buzzi Unicem Orlus 2007 (to I.D.), CIFE (to L.D.), AIRC (to L.D., D.U., S.B) and Human Genetics Foundation - HUGeF (to G.M.). The Turin case-control study was supported by a grant from Regione Piemonte. Ricerca Scientifica Applicata 2003 (to D.M.). The Casale case-control study was supported by a grant from Regione Piemonte. Ricerca Sanitaria Finalizzata 2004 (to C.M.). The Australian studies have been supported by the Australian National Health and Medical Research Council, the Sir Charles Gairdner Hospital and PathWest laboratory Medicine of WA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declare no competing financial interest. In fact, the “PathWest Laboratory Medicine WA” is not a commercial funder of this research. The authors Jennie Hui and John Beilby are employed by PathWest and do not have any additional consultancy, patents, products in development or marketed products with competing interests relating to this research, Thus, PathWest affiliation does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: giuseppe.matullo@unito.it (GM); irma.dianzani@med.unipmn.it (ID)
Introduction

Malignant pleural mesothelioma (MPM) is a rare, aggressive tumor that generally causes death within 2 years. The only clearly established risk factors for MPM are asbestos exposure, and exposure to erionite, other mineral fibers and x-ray for medical purposes [1]. Asbestos fibers retained in the lung and pleura may be carcinogenic, either through direct mechanical or biochemical effects, or through the activation of inflammatory cells. Persistent inflammation can induce chronic oxidative stress, genotoxic lesions, chromosomal aberrations and epigenetic alterations [2,3]. Asbestos fibers may also interfere with chromosome segregation and mitosis [4].

Although asbestos has been banned in many Western countries, it is still used in several parts of the world, and some developing countries are actually increasing the industrial use of asbestos, as well as its production and importation [5,6,7]. In Western Europe, over 5,000 people with MPM die each year [8,9,10,11]. The general characteristics of the Italian study sample, after quality controls (QC), are reported in Table 1 (392 MPM cases and 367 controls; 540 males, 219 females). A total of 330,879 SNPs were included in the analyses. The principal component analysis (PCA) (Figure S1) showed population stratification with two distinct clusters, which was further confirmed by K-mean analysis (data not shown). After correction of the regression analyses by PCA-cluster, the lambda inflation factor was <1.03 for both the overall and the exposed-only samples (Quantile- Quantile, QQ plots, Figure S2). Manhattan plots of the two-sided logistic regression analyses (per allele additive model) are also reported (Figure 1).

Results

The general characteristics of the Italian study sample, after quality controls (QC), are reported in Table 1 (392 MPM cases and 367 controls; 540 males, 219 females). A total of 330,879 SNPs were included in the analyses. The principal component analysis (PCA) (Figure S1) showed population stratification with two distinct clusters, which was further confirmed by K-mean analysis (data not shown). After correction of the regression analyses by PCA-cluster, the lambda inflation factor was <1.03 for both the overall and the exposed-only samples (Quantile- Quantile, QQ plots, Figure S2). Manhattan plots of the two-sided logistic regression analyses (per allele additive model) are also reported (Figure 1).

The genotyped SNPs with the highest significance levels are listed in Table 2. The imputed SNPs with the highest significance levels are listed in Table S1. Nine intragenic SNPs (7 genotyped and 2 imputed) were located in genes. When analyzing these nine genes in a Gene Set Enrichment Analysis (GSEA, File S1), significant enrichment involving MMP14 and ADAMTS2 was shown for gene-ontology (GO, File S1) biological processes including lung development (P=0.0087), respiratory tube development (P=0.0087), respiratory system development (P=0.0087), metalloendopeptidase activity (P=0.0140), and metallopeptidase activity (P=0.0210) (Table S2).

When the GSEA (File S1) was extended to SNPs with a significance level of P≤10^-3 in the regression analysis (additive model, 201 genes), another metallopeptidase, namely MMP9, was included in the gene list, further reinforcing the putative role of the metalloendopeptidase pathway in MPM.

Haplotype association was investigated in the Italian study sample for the 20 genes/chromosomal regions with the highest significance levels. The most significant haplotype associations were found in the chromosomal region 3p24.2, where the THRB...
gene is located \( P = 2.04 \times 10^{-7} \), and in 19q13.42 \( P = 7.02 \times 10^{-7} \) (Table S3), strengthening the importance of these chromosomal regions.

Seven chromosomal regions were significantly associated with MPM in the region-based analysis \( P < 0.0025 \), Table 3, Figure 2, Figure S3) [20]. The gene-based analysis confirmed the significance of the \( THRB \) gene \( P = 2.29 \times 10^{-5} \) and showed a borderline significance for the \( PVT1 \) gene \( P = 0.02 \) (Table 3). Finally, the regional GO (File S1) process-based analysis supported the involvement of the metalloendopeptidase and metallopeptidase GO (File S1) processes \( P = 0.0005 \) and 0.0039, respectively).

We detected a substantial improvement in accuracy comparing the first multivariate model, which used asbestos exposure as a predictor and adjusted for demographic covariates, with the second one, which also included 10 selected SNPs with independent effects (Table 4). The average Akaike Information Criterion (AIC) and area under ROC curve (AUC) across 10,000 random splits of the entire Italian study sample were 871.34 and 0.76 for the first model, and 730.27 and 0.86 for the second model, respectively (Figure 3, Table 4). The analysis stratified by center (Casale Monferrato versus Turin-Genoa) confirmed the stability of the risk estimates and 95% CIs (data not shown).

The first multivariate model confirmed asbestos exposure as the main risk factor for MPM (high exposure: OR 17.33, 95% CI 9.28–32.37, \( P < 2 \times 10^{-16} \); low exposure: OR 8.01, 95% CI 4.41–14.54, \( P = 8.52 \times 10^{-12} \) (Table 4). The second model, which included the genetic component, showed that the 10 selected SNPs had an independent contribution to MPM risk (Table 4), and also increased the estimate for the effect of asbestos exposure (high

---

**Figure 1. Manhattan plot of genotyped SNPs from logistic additive model.** A) all samples, B) exposed samples.
doi:10.1371/journal.pone.0061253.g001
### Table 2. Italian top 12 genotyped SNP list (2-tailed logistic regression, n = 759 overall, n = 593 exposed only).

<table>
<thead>
<tr>
<th>CHR Location</th>
<th>SNP</th>
<th>Ref. Allele</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>Typed</th>
<th>Gene Name</th>
<th>Left Gene</th>
<th>Right Gene</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>6q21</td>
<td>rs742109</td>
<td>A</td>
<td>0.55 (0.43–0.71)</td>
<td>2.70×10⁻⁶</td>
<td>Genotyped</td>
<td>PRDM1</td>
<td>ATG5</td>
<td>OVERALL</td>
<td></td>
</tr>
<tr>
<td>3q26.2</td>
<td>rs7632718</td>
<td>A</td>
<td>1.83 (1.42–2.37)</td>
<td>3.71×10⁻⁶</td>
<td>Genotyped</td>
<td>SLCA14, CLDN11</td>
<td>CLDN11</td>
<td>RPL22L1</td>
<td>EXPOSED</td>
</tr>
<tr>
<td>3p24.2</td>
<td>rs9833191</td>
<td>C</td>
<td>0.54 (0.41–0.71)</td>
<td>7.67×10⁻⁶</td>
<td>Genotyped</td>
<td>THRB</td>
<td>NR1D2</td>
<td>MIR4792</td>
<td>EXPOSED</td>
</tr>
<tr>
<td>5q23.1</td>
<td>rs1508805</td>
<td>A</td>
<td>1.85 (1.41–2.44)</td>
<td>1.04×10⁻⁵</td>
<td>Genotyped</td>
<td>PRR16</td>
<td>FTMT</td>
<td>EXPOSED</td>
<td></td>
</tr>
<tr>
<td>1q25.2</td>
<td>rs2501618</td>
<td>A</td>
<td>2.18 (1.53–3.10)</td>
<td>1.49×10⁻⁵</td>
<td>Genotyped</td>
<td>CEP350</td>
<td>TOR1AIP1</td>
<td>QSOX1</td>
<td>EXPOSED</td>
</tr>
<tr>
<td>5q35.3</td>
<td>rs4701085</td>
<td>G</td>
<td>1.84 (1.39–2.44)</td>
<td>1.93×10⁻⁵</td>
<td>Genotyped</td>
<td>ADAMT52</td>
<td>ZNF354C</td>
<td>AX747985</td>
<td>EXPOSED</td>
</tr>
<tr>
<td>4q22.1</td>
<td>rs4290865</td>
<td>A</td>
<td>1.98 (1.44–2.71)</td>
<td>2.16×10⁻⁵</td>
<td>Genotyped</td>
<td>FAM190A</td>
<td>GRID2</td>
<td>OVERALL</td>
<td></td>
</tr>
<tr>
<td>13q14.3</td>
<td>rs9536579</td>
<td>A</td>
<td>0.54 (0.40–0.72)</td>
<td>2.33×10⁻⁵</td>
<td>Genotyped</td>
<td>OLFM5</td>
<td>MIR1297</td>
<td>OVERALL</td>
<td></td>
</tr>
<tr>
<td>7p21.2</td>
<td>rs3801094</td>
<td>A</td>
<td>1.75 (1.35–2.27)</td>
<td>2.52×10⁻⁵</td>
<td>Genotyped</td>
<td>ETVI</td>
<td>AR14A</td>
<td>DGX8</td>
<td>OVERALL</td>
</tr>
<tr>
<td>8q24.21</td>
<td>rs7841347</td>
<td>A</td>
<td>0.60 (0.47–0.76)</td>
<td>2.60×10⁻⁵</td>
<td>Genotyped</td>
<td>PVT1</td>
<td>MYC</td>
<td>TME75</td>
<td>OVERALL</td>
</tr>
<tr>
<td>15q21.1</td>
<td>rs10519201</td>
<td>A</td>
<td>2.36 (1.57–3.56)</td>
<td>3.82×10⁻⁵</td>
<td>Genotyped</td>
<td>SHC4</td>
<td>EID1</td>
<td>SECISBP2L</td>
<td>EXPOSED</td>
</tr>
<tr>
<td>22q12.3</td>
<td>rs5756444</td>
<td>G</td>
<td>0.60 (0.47–0.76)</td>
<td>3.95×10⁻⁵</td>
<td>Genotyped</td>
<td>CSF2RB2</td>
<td>C22orf33/TEX33</td>
<td>EXPOSED</td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0061253.t002

### Table 3. Region-, Gene- and GO process-based analysis on top SNPs (1-tailed binomial test, n = 759, alpha 0.0025, alpha = 0.01, alpha = 0.025, respectively).

<table>
<thead>
<tr>
<th>Region/Gene/GO processes based</th>
<th>Cytogenetic Band</th>
<th>Position (from - to)</th>
<th>Number of SNPs</th>
<th>Significant SNPs</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1q25.2</td>
<td>(178192161–178267165)</td>
<td>5</td>
<td>4</td>
<td>8.31×10⁻⁴</td>
</tr>
<tr>
<td>-</td>
<td>3p24.2</td>
<td>(24311166–24397755)</td>
<td>17</td>
<td>7</td>
<td>3.86×10⁻⁴</td>
</tr>
<tr>
<td>-</td>
<td>3q26.2</td>
<td>(171668688–171738200)</td>
<td>12</td>
<td>6</td>
<td>9.47×10⁻⁵</td>
</tr>
<tr>
<td>-</td>
<td>4q22.1</td>
<td>(92842088–92925574)</td>
<td>11</td>
<td>3</td>
<td>0.05</td>
</tr>
<tr>
<td>-</td>
<td>4q32.1</td>
<td>(160680345–160763147)</td>
<td>11</td>
<td>3</td>
<td>0.04</td>
</tr>
<tr>
<td>-</td>
<td>5q23.1</td>
<td>(120950796–121034917)</td>
<td>11</td>
<td>3</td>
<td>0.08</td>
</tr>
<tr>
<td>-</td>
<td>5q35.2</td>
<td>(173515657–173599925)</td>
<td>16</td>
<td>4</td>
<td>7.23×10⁻³</td>
</tr>
<tr>
<td>-</td>
<td>5q35.3</td>
<td>(178559043–178654962)</td>
<td>19</td>
<td>5</td>
<td>0.01</td>
</tr>
<tr>
<td>-</td>
<td>6q21</td>
<td>(106656091–106738553)</td>
<td>18</td>
<td>5</td>
<td>8.00×10⁻³</td>
</tr>
<tr>
<td>-</td>
<td>7p21.2</td>
<td>(13877273–13974190)</td>
<td>20</td>
<td>6</td>
<td>4.36×10⁻³</td>
</tr>
<tr>
<td>-</td>
<td>7p22.2</td>
<td>(4339181–4436371)</td>
<td>17</td>
<td>9</td>
<td>5.96×10⁻⁵</td>
</tr>
<tr>
<td>-</td>
<td>8q24.21</td>
<td>(128837336–128953399)</td>
<td>7</td>
<td>6</td>
<td>1.04×10⁻⁴</td>
</tr>
<tr>
<td>-</td>
<td>9p24.1</td>
<td>(5363441–5453988)</td>
<td>12</td>
<td>5</td>
<td>0.02</td>
</tr>
<tr>
<td>-</td>
<td>12q23.3</td>
<td>(107375486–107461372)</td>
<td>13</td>
<td>7</td>
<td>5.78×10⁻⁵</td>
</tr>
<tr>
<td>-</td>
<td>13q14.3</td>
<td>(53429288–53513774)</td>
<td>12</td>
<td>4</td>
<td>0.02</td>
</tr>
<tr>
<td>-</td>
<td>14q11.2</td>
<td>(22334110–22425388)</td>
<td>13</td>
<td>2</td>
<td>0.14</td>
</tr>
<tr>
<td>-</td>
<td>15q14</td>
<td>(34381353–34470568)</td>
<td>13</td>
<td>5</td>
<td>2.04×10⁻³</td>
</tr>
<tr>
<td>-</td>
<td>15q21.1</td>
<td>(46959609–47047893)</td>
<td>18</td>
<td>2</td>
<td>0.23</td>
</tr>
<tr>
<td>-</td>
<td>19q13.42</td>
<td>(59189856–59266559)</td>
<td>9</td>
<td>1</td>
<td>0.47</td>
</tr>
<tr>
<td>-</td>
<td>22q12.3</td>
<td>(35600028–35754794)</td>
<td>19</td>
<td>5</td>
<td>0.03</td>
</tr>
<tr>
<td>CEP350</td>
<td>1q25.2</td>
<td>(179933906–180093734)</td>
<td>17</td>
<td>2</td>
<td>0.31</td>
</tr>
<tr>
<td>THR8</td>
<td>3p24.2</td>
<td>(24162088–24541232)</td>
<td>54</td>
<td>15</td>
<td>2.29×10⁻³</td>
</tr>
<tr>
<td>SLCA14</td>
<td>3q26.2</td>
<td>(170167538–1711715102)</td>
<td>13</td>
<td>2</td>
<td>0.16</td>
</tr>
<tr>
<td>SDK1</td>
<td>7p22.2</td>
<td>(3341374–4303003)</td>
<td>90</td>
<td>5</td>
<td>0.61</td>
</tr>
<tr>
<td>PVT1</td>
<td>8q24.21</td>
<td>(128808953–129119976)</td>
<td>34</td>
<td>7</td>
<td>0.02</td>
</tr>
<tr>
<td>METALLOENDOPEPTIDASE</td>
<td>-</td>
<td>-</td>
<td>197</td>
<td>19</td>
<td>4.65×10⁻³</td>
</tr>
<tr>
<td>METALLOPEPTIDASE</td>
<td>-</td>
<td>-</td>
<td>470</td>
<td>32</td>
<td>0.04</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0061253.t003
exposure: OR 45.28, 95% CI 21.52–95.28, \( P = 2 \times 10^{-16} \); low exposure: OR 15.31, 95% CI 7.78–30.14, \( P = 2 \times 10^{-15} \)).

**SNP validation and replication**

The Italian and Australian study samples showed a marked degree of heterogeneity (I² statistics, range 0.62–0.97) [21] (Table S5). None of the 12 genotyped SNPs with the highest significance levels in the Italian study were found in the Australian replication study (Table S4), and nor of these were confirmed by the meta-analysis (Table S5). Nevertheless, when a regional analysis was performed in the Australian study sample, we found significant associations in five chromosomal regions (3q26.2, 4q32.1, 7p22.2, 14q11.2, 15q14) that have reported to be altered in mesothelioma (Table 5) [20].

**Gene expression analysis in blood and in normal pleural tissue**

Gene expression analysis on lymphocytes from Italian healthy subjects (Text S1) showed a possible expression Quantitative Trait Locus (eQTL) for the \( PVT1 \) (rs7841347) gene (non-parametric Kruskal-Wallis test \( P = 0.001 \)) (Figure 4). However, expression analysis from Italian healthy subjects pleural tissue stratified by \( PVT1 \) rs7841347 genotypes did not show any gradient, although a statistically significant difference \( (P = 0.01) \) was found (Figure S4). Published expression data [22] (Text S1) confirmed the dysregulation of \( MMP14 \), \( THRB \) and \( MYC \) genes in MPM, supporting our results.

**SNP predictive functional analysis**

Using the GenomePipe tool, none of the SNPs with the highest significance levels included in the present analysis might predict damage, nor were they located in a regulatory or splicing site.

---

**Figure 2. Regional association plots for 4 of the most consistent chromosome regions.** a. 3p24.2, b. 8q24.21, c. 14q11.2, d. 7p22.2. Consistency was based on haplotype, gene-, region- and pathway analysis. Each SNP is plotted with respect to its chromosomal location (x axis) and its log\(_{10}\) transformed \( P \) value (y axis on the left) for associations with MPM. The tall blue spikes indicate the recombination rate (y axis on the right) at that region of the chromosome. The red-outlined diamond indicate the index SNP and other diamond indicate the genotyped SNPs, the squares indicate imputed SNPs using as reference 1000 Genomes Pilot 1 CEU population. LD values were calculated only on our control population.

doi:10.1371/journal.pone.0061253.g002
Even when SNPs in Linkage Disequilibrium (LD) with our top SNPs (LD $r^2 < 0.8$ as measured by pairwise $r^2$) were included in the analysis no evidence of functional properties of the proxy SNPs was found. LD refers to two different populations, i.e. HapMap TSI from Tuscany (Italy) and CEU (HapMap3, File S1), for a total of 33 and 72 SNPs respectively.

**Discussion**

In order to identify genetic risk factors that might contribute to the development of MPM, we performed a GWAS on 407 Italian MPM cases and 389 controls.

We performed an independent replication study in an Australian sample, which included 428 MPM cases (Genetic Understanding of Asbestos-Related Disease, GUARD, study) and 1,269 controls (Busselton Health Study, BHS).

Among the top SNPs identified in our Italian study sample, there were several genes previously reported to be involved in MPM or other cancer types, as well as chromosomal regions reported to be altered in MPM [20].

Although no single SNP replicated in the Australian sample, probably due to the high genetic heterogeneity between the two studies, regional analyses showed significant signals in 5 of the chromosomal regions where the Italian top SNPs are located. The chromosomal region 7p22.2 found in the replication study includes the $SDK1$ [23] and $FOXK1$ [24] genes. Interestingly, $FOXK1$ has been reported to interact with $BAP1$ [25], which was recently found to be mutated in mesothelioma [19]. Chromosomal region 7p22 is located in a fragile sequence (FRA7B) containing two miRNA genes (mir589 and mir339) and three large genes ($SDK1$, $THSD7A$, $MAD1L1$), and is highly prone to gaps and breaks in several cancers [23].

Another Italian genotyped top-signal (rs7632718) is located in the $SLC7A14$ (solute carrier family 7 member 14) gene, which lies on 3q26.2, which was one of the replicating regions in the Australian study. Although no link with MPM has been previously reported for $SLC7A14$, a chromosomal gain has been described in this region [20], suggesting a possible involvement of other genes in MPM.

The $PVT1$ (Pvt1 oncogene (non-protein coding)) gene is involved in several types of cancer [26,27,28,29,30]. It is located in a large (>300 kb) locus downstream of $MYC$ (53 Kb apart) on chromosomal region 8q24. The $PVT1$ locus produces a wide variety of spliced non-coding RNAs as well as a cluster of six annotated miRNAs: miR-1204, miR-1205, miR-1206, miR-1207-5p, miR-1207-3p, and miR-1208 [31,32]. $PVT1$ was proposed to regulate $c$-Myc gene transcription over a long distance [33]. A functional variant (rs378854) in chromosomal region 8q24 that modulates $PVT1$ expression has been associated with prostate

**Table 4.** Nested multivariate logistic regression models: 1) model 1, without genetic component; 2) model 2, with genetic component.

<table>
<thead>
<tr>
<th>MODEL 1</th>
<th>MODEL 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>OR_95L</td>
</tr>
<tr>
<td>LOW vs NO EXPOSURE</td>
<td>8.01</td>
</tr>
<tr>
<td>HIGH vs NO EXPOSURE</td>
<td>17.33</td>
</tr>
<tr>
<td>CLUSTER 2 vs 1</td>
<td>1.76</td>
</tr>
<tr>
<td>rs2501618</td>
<td>-</td>
</tr>
<tr>
<td>rs9833191</td>
<td>-</td>
</tr>
<tr>
<td>rs7632718</td>
<td>-</td>
</tr>
<tr>
<td>rs4701085</td>
<td>-</td>
</tr>
<tr>
<td>rs73034881</td>
<td>-</td>
</tr>
<tr>
<td>rs3801094</td>
<td>-</td>
</tr>
<tr>
<td>rs7841347</td>
<td>-</td>
</tr>
<tr>
<td>rs10815216</td>
<td>-</td>
</tr>
<tr>
<td>rs2236304</td>
<td>-</td>
</tr>
<tr>
<td>rs7178364</td>
<td>-</td>
</tr>
</tbody>
</table>

*adjusted for age, gender and center of recruitment.

MODEL 1: AIC = 871.3, AUC = 0.76.
MODEL 2: AIC = 730.27, AUC = 0.86.
doi:10.1371/journal.pone.0061253.t004

![Figure 3. Receiver Operating Curves (ROC) for the two multivariate models including asbestos exposure 1) without and 2) with the 10 most robust and significant genetic variants. doi:10.1371/journal.pone.0061253.g003](image-url)
In vitro, the rs378854-G allele has been associated with reduced binding of the transcription factor YY1, a putative tumor suppressor, and with repressed global transcription in prostate cancer [33]. The regulation of this chromosomal region is very complex, as is suggested by the association of several SNPs with different cancer types [35], and involves miRNA, lincRNA and other epigenetic regulations [36].

The gene-expression analysis on lymphocytes from Italian healthy subjects showed a possible eQTL for PVT1. Functional studies are needed to clarify the link between PVT1-associated SNPs, gene expression regulation and cancer risk taking into account that in our study PVT1 seems to act only at an early stage of carcinogenesis as its deregulation has not been observed at later stages in tumor tissue [22].

Two other genes that have been reported to be dysregulated in MPM, are THRB and MMP14 [22,37]. THRB encodes for thyroid hormone receptor beta (TRβ), which could function as a tumor suppressor. Cell-based studies and xenograft models have demonstrated that TRβ is a suppressor of ras-mediated cell proliferation, transformation, and tumorigenesis [38]. Moreover, TRβ disrupts mitogenic growth factors by suppressing the activation of extracellular signal-regulated kinases and phosphatidylinositol 3-kinase signaling pathways to suppress tumor cell invasiveness and metastasis [39,40]. THRB is located about 28 Mb telomeric to the BAP1 gene, which is mutated in MPM [19]. A down-regulation of
THRβ has been documented in MPM versus parietal pleura [41] and it is frequently methylated/deleted in non-squamous-cell lung cancer [42].

MMP14 (matrix metallopeptidase 14) has been reported to influence overall survival in MPM cases [37], and was significantly highlighted in our enrichment analysis, together with ADAMTS2, because of their metalloendopeptidase and metallopeptidase activities. The matrix metalloproteinases are a family of zinc-containing enzymes with proteolytic activity against a wide range of extracellular proteins. Extracellular matrix proteases are involved in several steps of cancer development and progression, including angiogenesis and metastasis.

Some of the SNPs with highest significance levels were located in the genes: CEP350, ETV1 and SHCA. Although they have not been directly associated with MPM, their involvement in several cancer types has been described [43,44,45], suggesting the necessity to further investigate their possible role in MPM pathogenesis. Considering the closest flanking genes of intergenic SNPs, the following are noteworthy and could contribute to the carcinogenic process, as has been reported for other cancer types: PRDM1 [46], ATG5 [47], MTC [48], EID [49], RLN1 [50], CD274 [51].

Although our sample size is clearly a limitation for a GWAS, the Italian and the Australian study samples are, to the best of our knowledge, the largest MPM series with available DNA, as mesothelioma is a very rare cancer. A further limitation of GWAS is that they do not take into account rare variants. The availability of methods for complete genome sequencing (and the decrease of the sequencing costs) will allow to circumvent the problem linked to the identification of rare variants, whose involvement should be better investigated in future studies.

The negative replication of the Italian top SNPs in the Australian study should be revised on the basis of the following considerations: i) the two studies had a marked degree of heterogeneity as shown by the I² statistics; ii) no exposure assessment was available for the Australian control group. Notwithstanding these discrepancies, we observed an intriguing significant regional replication in the Australian study for 5 out of 20 Italian top signals.

Most of the top-signals we identified were located in chromosomal regions reported to harbor aberrant alterations in mesothelioma, and cause an at most 2–3 fold increase in MPM risk.

Moreover, asbestos exposure in our study was associated with a remarkable increase in MPM risk, which became even more evident when the contribution of genetic factors was taken into account, with a significant improvement of asbestos exposure risk estimation.

In conclusion, our results support the complementary role of genetic background in asbestos-related carcinogenesis of the pleura, indicating that genetic risk factors should be taken into account to understand MPM physiopathology, and to better define the MPM risk profile of people with a high exposure to asbestos.

Methods

Ethics statement

All MPM cases reported on in the present report gave written informed consent. This study was performed according to the principles of the Declaration of Helsinki and in agreement with ethical requirements. Approval was obtained from the Istituto Nazionale per la Ricerca sul Cancro Ethics Committee for the studies in Genoa and La Spezia, and from the Human Genetics Foundation (HuGeF) Ethics Committee for the studies in Casale Monferrato and Turin. The Australian replication study was specifically approved by the Human Research Ethics Committee of the University of Western Australia.

Italian study sample

The Italian study sample is composed of MPM cases and controls from cities located in Northern Italy: Casale Monferrato and Turin in the Piedmont Region, and Genoa and La Spezia in the Liguria Region (Table 1; details in Text S1). The study in Casale Monferrato was a population-based MPM case-control study [52], and included 241 MPM patients and 252 population controls of Italian nationality and Caucasian ethnicity. The study in Turin was a hospital-based MPM case-control study [11], and consisted of 91 MPM patients and 56 controls of Italian nationality and Caucasian ethnicity. The hospital-based study in Genoa and La Spezia included 75 incident MPM cases [53]. Controls are 81 healthy subjects or patients hospitalized for non-neoplastic/non-respiratory conditions.

All the three of the above-mentioned Italian studies were registry-based and therefore no selection criteria were applied to MPM cases; they needed only to be residing in the study area at the time of diagnosis. Only cases with a pathological diagnosis (based on histology or cytology with confirmatory immunohisto-chemical staining) were eligible for inclusion in the present analysis. Study periods in the Italian studies were different (Casale Monferrato: January 2001 to December 2006; Turin: January 2004 to October 2008; Genoa and La Spezia: April 1996 to February 2006 for cases and February 1997 and November 2006 for controls). For practical reasons, the study in Turin was limited to cases admitted to the main metropolitan hospitals.

Asbestos exposure was carefully assessed in all the Italian cases and controls. After reviewing individual occupational histories, asbestos exposure was reclassified for the overall sample by the same expert (D.M.) as “no/unlikely” (no acknowledged occupational or environmental exposure), “low” (low exposure probability, or definite low exposure), and “high” (definite high exposure; asbestos-cement and asbestos-textile workers, insulators, shipyard workers and dockers).

Australian replication study

Australian MPM cases were part of the GUARD study, which consisted of individuals who had been exposed to asbestos and diagnosed with MPM (n = 428) and who attended a hospital clinic in Perth, Western Australia between 1988 and 2010 [54]. DNA samples and clinical data from these individuals were obtained and MPM diagnosis was confirmed after pathological, radiological and clinical review with confirmation from respective cancer registries in Western Australia (Western Australia Mesothelioma Registry) and Queensland.

The GUARD study subjects are primarily male (88.8%) with an average age of 67±10.3 years. Most BHS study subjects are female (57.4%) and the average age is 54±17.2 years. Control samples (n = 1,269), with no information on asbestos exposure, were obtained from the population-based BHS [53]. MPM cases were excluded after genotyping if they were: related to another individual, had a low call GWAS rate (<97%), were not Caucasian/European based on principal component analysis, had ambiguous sex, or had low heterozygosity compared to the rest of the sample.

SNP genotyping

Whole-genome genotyping was done on a HumanCNV370-Quad BeadChip (Illumina Inc., San Diego, CA, USA) for 716 samples. The remaining 80 samples were tested on a Human610-
Quad (which includes 100% of the HumanCNV370 BeadChip SNPs) as the HumanCNV370-Quad had been discontinued. Genotypes were assessed by GenomeStudio V2011.1 (Illumina Inc., San Diego, CA). The 12 most significant SNPs from the Italian study S were individually genotyped in the Australian replication study with a 5’-nuclease assay (Applied Biosystems, CA, USA).

Statistical analysis

**Genotyping quality controls.** A cut-off a genotyping call rate of 0.98 was set, leading to the exclusion of 18 study subjects. SIdentity By Descent (IBD) estimation using the Identity By State (IBS) distance was used to check genotypic identity or relatedness among subjects (PLINK software [56], File S1). Subjects with IBD=0.05 (n=16) were consideredconsanguineous and excluded from further analyses. We additionally excluded three samples with an X chromosome inbreeding homozygosity estimate of about 0.5. Thirty-seven subjects (4.6%) were removed from the analysis, leaving 759 subjects (392 cases and 367 controls).

SNPs with minor allele frequency $<1%$ (n=13,252), those having $>0.05$ missing genotypes (n=11,535) and those deviating from Hardy-Weinberg equilibrium (HWE) in the control population (P=$<0.001$, n=1,157) were excluded from the analysis, for a final study data-set of 330,879 SNPs, which were analyzed for their potential association with mesothelioma.

**Population structure and association analysis.** The population structure was investigated by PCA (PLINK Software, File S1, Covariance Method [57]). A new discrete covariate was defined by the two principal components (Figure S1), and was included in the following logistic regression analysis. PCA results were further confirmed by the K-means clustering analysis [58] (data not shown). The effective removal of any population structure bias was checked by the $\lambda$-inflation factor parameter [59] (Figure S2).

We tested for 330,879 SNPs for their association with mesothelioma by 2-sided logistic regression analysis on a per-allele additive model after adjusting for age, gender, PCA cluster, center of recruitment and exposure level, both in the overall Italian sample (n=759) and among exposed-only Italian subjects (n=593) [60]. After Bonferroni correction, we considered alpha $=1.51 \times 10^{-7}$ (0.05/330879) as a threshold of significance. The analyses were performed with PLINK v1.07 (File S1) [56] and Rv2.10.1 [60] software. The software Impute.v2 [61,62] was used to impute 5,333,982 SNPs, using the 1000 genomes (http://www.1000genomes.org/) and HapMap3 (File S1) genotype panels as reference datasets.

Haplotypes (Table S3) within the chromosomal regions where the top significant SNPs were located (considering sliding windows from 2 to 10 SNPs; PLINK Software, File S1) were also tested for any association with MPM in the overall Italian sample.

**Meta-analysis and replication.** A meta-analysis of the Italian-study top 12 genotyped SNPs was done on data from the whole genome genotyping (Human610-Quad BeadChip, Illumina) of 428 cases and 1269 Australian controls of European descent (GWAMA software, File S1 [63]). A random-effects model was used due to the presence of genetic heterogeneity ($I^2$ statistic [21] $>50%$; Table S3).

**Multivariate analysis.** The cumulative effect of the SNPs with highest significance levels was investigated by two-sided multivariate logistic regression analysis, comparing the prediction accuracy of two models: the first considering asbestos exposure as a predictor and adjusting for demographic covariates (recruitment center, gender, age, geographical cluster), and the second identical to the first, but also including the genetic component (genotypes). SNPs included in the second multivariate model were selected among the top 20 markers (12 genotyped and 8 imputed), excluding 4 SNPs (rs4290865, rs1354252, rs1072577, rs10519921) because of negative internal replication between Casale Monferrato and pooled Turin-Genova studies, and 6 SNPs (rs742109, rs1308805, rs9356579, rs576444, rs6897549, rs71365421) because they did not replicate in the Australian study on the regional analysis and were not intragenic.

An internal validation of the two models was done by randomly splitting the overall Italian sample in two groups 10,000 times, each time performing a two-sided logistic regression in the first group and verifying the accuracy of estimation in the second group. The average AIC under 10,000 permutations and AUC were used as measures of the fit and the prediction power of the two models.

**Gene-region enrichment and SNP functional prediction analyses.** A GSEA (File S1) [64] was performed on the genes in which the top SNPs are located (9 genes out of 20 signals: PTIT1 [gene ID 5820], CEP550 [ID 9507], THRB [ID 7068], ETT1 [ID 2115], C9orf46 (also known as PLGRK7, ID 55506), MMP14 [ID 439], ADAMTS2 [ID 9509], SLCTA14 [ID 57709], SLC4A (ID 399694). The list was tested for over-representation using the curated Molecular Signatures Database (MSigDB) 7, specifically i) KEGG 0 (File S1), REACTOME and BioCarta pathway databases, ii) the GO (File S1) gene set 9. Gene set enrichment significance was tested by a hyper-geometric test that evaluates the distribution of overlapping genes over all genes in the gene set (Table S2).

Region-, gene- and GO (File S1) process-based analyses were also performed [65]. We investigated the occurrence of multiple signals in those genes and chromosomal regions, where the significant SNPs from the single SNP analysis are located, as well as those from genes belonging to the pathways identified by the GO (File S1) process-based analysis (Table 3).

We tested 20 candidate chromosomal regions, and five genes (CEP550, THRB, SLC7A14, SDK1 and PTIT1) for which there were enough representative SNPs genotyped, and two GSEA significant GO processes (File S1) (metalloendopeptidase activity and metalloendopeptidase activity). After Bonferroni correction, we adopted the following significance thresholds: alpha $=0.025$, alpha $=0.01$, alpha $=0.025$, for region-based, gene-based and GO (File S1) process-based analysis respectively.

Prediction of functional SNPs has been carried out with several softwares, including GenomPipe software, which is freely available at website of the National Institute of Environmental Health Sciences (http://snpinfo.niehs.nih.gov/selGWAs.htm) and the Pupasuite3.1 software (http://pupasuite.bioinfo.icp.es/).

**Gene-expression analysis.** The expression levels of the nine genes corresponding to the most common intragenic SNPs (Table 2) and of MYC, which is neighbor to PTIT1, were examined using data from the HapMap (File S1) CEU gene-expression database, and the GenoPheno database [66], an internal database which includes genotypic, phenotypic, and gene-expression data from the peripheral blood of 120 healthy Italian volunteers (Text S1). We considered the average expression levels of probes and, when feasible, tested for differential expression among the three genotypes (Kruskal-Wallis test).

In addition, the mRNA levels of the PTIT1, MYC and THRB genes were measured by quantitative real-time PCR in 79 normal pleural tissues from donors that underwent thoracoscopy for conditions other than MPM, who signed an informed consent form (Text S1).
Supporting Information

Figure S1 Principal Component Analysis (PCA) plots: first vs second PC. A) Cases and controls are plotted for the overall study and for each of the three study samples (Turin, Casale Monferrato and Genoa); B) birth places (Northern, Central, Southern Italy, Sardinians and Other Caucasians) are plotted for the overall study and for each of the three study samples. (TIFF)

Figure S2 Supplementary figure 1: Q-Q plots for GWAS of mesothelioma in the Italian population. This Q-Q plots are based on logistic regression allele P after standard quality control. The estimated λ inflation factor was <1.03. Plot A shows the Q-Q plot for the overall Italian population, whereas Plot B refers to the exposed-only population. (TIFF)

Figure S3 Regional association plots for additional 4 regions (a. 3q26.2, b. 4q32.1, c. 7p21.2, d. 15q14) replicating in the Australian study. Each SNP is plotted with respect to its chromosomal location (x axis) and its log_{10} transformed P value (y axis on the left) for associations with MPM. The tall blue spikes indicate the recombination rate (y axis on the right) at that region of the chromosome. The red-outlined diamond indicate the index SNP and other diamond indicate the genotyped SNPs, the squares indicate imputed SNPs using as reference 1000 Genomes Pilot 1.

Table S1 Italian top 8 imputed SNP list. (DOCX)

Table S2 Gene Set Enrichment Analysis. (DOCX)

References

Table S3 Significant Haplotype Results for 3p24 and 19q13.42 regions. (DOCX)

Table S4 Replication of the 12 genotyped Italian top SNPs on GUARD-BHS Study. (DOCX)

Table S5 Meta-analysis of Italian and Australian studies for the top 12 genotyped Italian SNPs. (DOCX)

Text S1 Supplementary Materials. (DOCX)

File S1 URLs. (DOCX)

Acknowledgments
We wish to thank all the patients and healthy controls that generously participated in the study, the attending MDs that supported it, Dr Silvia Polidoro for her contribution to our the first technical issues with the Illumina platform, and Progetto RoPHS, Regione Piemonte - Bando Scienze umane e sociali (L.D.).

Author Contributions
Contributed to writing the manuscript: SG CDG AR GF FR. Evaluated asbestos exposure: DM. Obtained and/or supervised clinical information: RG EP AH FA ER CC SM. Obtained funding for sample collection and genotyping: ID GM CM LJP AJ SB DU DM PGB. Participated in critical review of the manuscript for intellectual content: GM SG MB GF DF FV GC CDG FR AR AH EC ST MP MG AA CC FA ER PGB RL RG EP MN AWBM NHkK JH JB ALJ JC BWR SM LJP DM DU SB CM ID. Conceived and designed the experiments: ID GM SB CM. Performed the experiments: SG AR MB AA EC FR ID GM JH. Analyzed the data: GF CMS DF ST FV GC CDG FR MP. Contributed reagents/materials/analysis tools: MB MN DU RL MG PGB SB AH JC BWR AWBM LJP ALJ NHkK JB CC FA. Wrote the paper: GM ID CM.

A GWAS on Malignant Pleural Mesothelioma
38. Garcia-Silva S, Aranda A (2004) The thyroid hormone receptor is a suppressor
24. Komorek J, Kuppuswamy M, Subramanian T, Vijayalingam S, Lomonosova E,
et al. (2010) The ubiquitin carboxyl hydroxylase BAP1 forms a ternary complex with YY1 and
HCF-1 and is a critical regulator of gene expression. Molecular and cellular
biology 30: 5071–5085.
technology in the study of mesothelioma. Journal of thoracic oncology: official
publication of the International Association for the Study of Lung Cancer 4: 411–425.
targets of malignant pleural mesothelioma. Mutation research.
common fragile site mapped at the 7p chromosome terminal region. Cancer
genetics and cytogenetics 202: 47–52.
11. Komorek J, Kuppuswamy M, Subramanian T, Vijayalingam S, Lomonosova E,
et al. (2010) Adenosine type 5 E1A and E6 proteins of low-risk cutaneous beta-
human papillomaviruses suppress cell transformation through interaction with
ubiquitin carboxyl hydroxylase BAP1 forms a ternary complex with YY1 and
HCF-1 and is a critical regulator of gene expression. Molecular and cellular
biology 30: 5071–5085.
Breakpoints of Burkitt’s lymphoma t(8;22) translocations map within a distance of
300 kb downstream of MYC. Genes, chromosomes & cancer 9: 202–207.
Aberrant ETV1 expression in acute myeloid leukemias associated with the
t(5;8)(p22;q24). Cancer genetics and cytogenetics 177: 37–42.
Identification of microRNA genes in genomically unstable regions of human
chromosome 8q24. Mol Cancer Res 6: 212–221.
ubiquitin carboxyl hydroxylase BAP1 forms a ternary complex with YY1 and
HCF-1 and is a critical regulator of gene expression. Molecular and cellular
biology 30: 5071–5085.