"This is the peer reviewed version of the following article: [ Johari, Mridul; Arumilli, Meharji; Palmio, Johanna; Savarese, Marco; Tasca, Giorgio; Massimiliano, Mirabella; Sandholm, Niina; Lohi, Hannes; Hackman, Peter; Udd, BjarneApr;24(4):572-577 Association study reveals novel risk loci for sporadic inclusion body myositis. Eur J Neurol. 2017 Apr;24(4):572-577], which has been published in final form at [Link to final article using the DOI]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**Association study reveals novel risk loci for Sporadic Inclusion Body Myositis**

Mridul Johari, MSc 1, Meharji Arumilli, MSc 1,2,3, Johanna Palmio, MD, PhD 4, Marco Savarese, PhD 1, Giorgio Tasca, MD 5, Mirabella Massimiliano, MD 6, Niina Sandholm, PhD 1,7,8, Hannes Lohi, PhD 1,2,3, Peter Hackman, PhD 1, Bjarne Udd, MD, PhD 1,4,9  

1Folkhälsan Institute of Genetics, Medicum, University of Helsinki, Helsinki, Finland  
2Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland  
3Department of Veterinary Biosciences, University of Helsinki, Finland  
4Neuromuscular Research Center, Tampere University and University Hospital, Tampere, Finland  
5Institute of Neurology, Policlinico “A.Gemelli” Foundation University Hospital, Rome, Italy  
6Institute of Neurology, Catholic University School of Medicine, Rome, Italy  
7Abdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland  
8Research Program Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland  
9Department of Neurology, Vaasa Central Hospital, Vaasa, Finland

**Corresponding author:**

Peter Hackman  
Folkhälsan Institute of Genetics  
Dpt. of Medical Genetics  
University of Helsinki  
Biomedicum, Haartmaninkatu 8 -Pb 63, FI-00014 Helsinki -Finland  
**E-Mail:** peter.hackman@helsinki.fi  
**Tel.** +358-294125069

Running head: Risk loci associated with sIBM  
Total word count: 3464  
Number of figures and tables: 1 figure and 2 tables

**Keywords:** Sporadic Inclusion Body Myositis (sIBM), Genetic risk factors, Risk loci, Whole Exome Sequencing, Association study, Case-Control study, Sphingolipids, HLA
Abstract

Background

To identify potential genetic risk factors associated with Sporadic Inclusion Body Myositis (sIBM).

Methods

We utilized an association based case-control approach on WES data of 30 Finnish sIBM patients and of a control cohort (n=193). A separate Italian cohort of sIBM patients (n=12) was used for evaluation of the results.

Results

We identified 7 SNPs in 5 genes that have an observed considerably higher frequency in Finnish sIBM patients compared to the control population, and can confirm the previous association of the genetic HLA region.

Conclusions

All seven identified variants could individually or in combination increase the susceptibility for sIBM.
Introduction

Sporadic inclusion body myositis (sIBM) is a complex, likely multifactorial disorder with an observed male predominance. It is the most common form of adult onset inflammatory myopathy in the world, with an age of onset after 45 years [1] and a reported prevalence ranging from 1 to 71 per million [2].

Characteristic clinical features of sIBM include distinct and often asymmetric slowly progressive muscle weakness and atrophy in deep finger flexors and distal knee extensors [3]. Hand grip and lower limb weakness is usually accompanied by dysphagia in approximately 65% of the cases [4] and loss of quality of life at later stages of the disease [5].

Recent studies emphasize the role of clinical features and histopathological findings in diagnosing sIBM [6, 7]. Muscle imaging is an additional diagnostic tool showing a consistent pattern of involvement [3, 8]. Muscle biopsy shows degeneration of fibers accompanied by inflammatory changes. These degenerative changes consist of atrophic fibers, rimmed vacuoles, congophilic inclusions, multi-protein aggregates and mitochondrial defects with COX-negative SDH-positive fibers[9, 10]. Endomysial lymphocyte infiltrates, predominantly CD8+ T-cells with invasion of non-necrotic muscle fibers, account for the inflammatory changes, and upregulation of class I MHC antigens is part of the immune response [11].

One recent hypothesis suggests that abnormal multi-protein aggregates cause the degeneration of aging muscle fibers [12]. Additionally, genetic immunological factors could increase susceptibility for the T-cell mediated inflammatory response [13].

The strongest reproduced association remains with the HLA antigens. *HLA-DR3* and the extended 8.1 ancestral haplotype have been strongly associated with increased susceptibility for sIBM [14]. 35.2 AH
and 52.1 AH along with other alleles and haplotypes have been associated with increased risk of sIBM in Caucasian and Japanese populations respectively [15-18].

Within the susceptibility MHC region, polymorphisms in the NOTCH4 gene have also been shown to have association with sIBM (OR > 2) in two independent Caucasian populations [19].

Recently an intronic polymorphism in the gene TOMM40 was suggested to have a disease-modifying effect [20]. A targeted sequencing study including 79 sIBM patients found rare VCP variants in two unrelated patients [21].

Taking advantage of the genetically more homogenous Finnish population background, we designed a WES based case-control study and observed an enrichment of certain specific variants in our sIBM population, underlining the role of genetic factors in the pathology of sIBM.
**Materials and Methods**

Thirty patients from Finland and 12 patients from Italy were included in the study. The age of onset of symptoms was over 40 years (mean=60.4 years) and family history was negative. The proportion of Finnish male patients to female patients was 60% male (18) and 40% female (12). Clinical phenotype and pathology were consistent with ENMC criteria and the diagnosis was classified as Clinico-pathologically defined IBM [7]. Muscle MRI, performed on 24 Finnish and 11 Italian sIBM patients, showed degenerative changes and edema in distal quadriceps femoris, and frequently in the medial gastrocnemius muscles [8]. In seven more advanced patients in the Finnish cohort, the degenerative findings were more widespread.

All the patients provided informed consent. The study was approved by the University of Helsinki ethics review board. Genomic DNA was extracted from blood leucocytes using standard procedure.

**Genetic and bioinformatic analysis**

WES for 30 Finnish sIBM patients was performed using the SureSelect v2 (Agilent Technologies, CA, USA). WES was performed for 94 controls using the SeqCap EZ Exome v2.0 (Roche Sequencing, CA, USA). Publicly available exome data for 99 Finnish controls was downloaded from 1000Genomes Project website (http://www.1000genomes.org). To counter the possible batch effects, only the common probes between all the batches were used for the association study.

In total, the study cohort included 30 Finnish sIBM patients and 193 control samples exome-sequenced on contemporary Illumina platforms.

For each sample, read alignment was done using BWA [22]. Duplicate reads were marked using Picard (https://github. Com/broadinstitute/picard) and local realignment was performed using GATK [23].
Joint genotyping was done on the cohort of 223 samples using GATK GenotypeGVCFs variant discovery tool.

**Candidate gene approach**

A candidate gene approach on WES data was used to investigate variants in 180 genes [24]. We filtered out all the common variants (minor allele frequency (MAF) higher than 1%) in the 1000genome, ExAC database and our internal control cohort.

**Association study**

Case-control variants were imported to GenABELv.1.6–7 [25] in R for genome-wide association analysis. The genotypes were filtered for MAF>0.01, genotype call rate>0.90, individual call-rate>0.95 and Hardy-Weinberg equilibrium p>0.0000001 in the controls. Multidimensional scaling plots were produced to identify outliers and population stratification. Identity-by-state values were calculated to control for duplicate samples and unknown relatedness. Mixed model approach was selected to correct for the relatedness between the genotyped individuals. The polygenic model was estimated using the hierarchical generalized linear model in GenABEL [26]. Score test was used for association testing. Association p-values were corrected for genomic inflation factor \( \lambda > 1 \) and adjusted by Bonferroni correction for multiple testing.

Human Splicing Finder (HSF) [27] was used to analyze the potential splicing effects.

Twelve Italian sIBM patients were used as a ‘validation cohort’ for the significant 7 identified SNPs. All significantly enriched SNPs were confirmed by Sanger sequencing.
Results

We analyzed the WES data of the 30 Finnish sIBM patients for rare variants in the extensive list of known myopathy disease genes [24]. No molecular findings able to explain the observed phenotype were identified, excluding the presence of a muscular disorder due to already known genes (Table 1).

In the following case-control association study we identified 7 SNPs with a nominal p-value<0.005 enriched in the Finnish sIBM population (Figure 1).

Two of the significantly enriched SNPs were identified in the HLA region. Rs1063318 in *HLA-DQB1* is a nonsynonymous change enriched 2.01 times in the sIBM case population. Similarly, rs7383287, a synonymous variant in *HLA-DOB*, was enriched 2.3 times.

Three SNPs (rs2941515, rs73296109 and rs2941513) were identified in the *STARD3* gene as a haplotype. All three of them were synonymous variants and were enriched 4 times. Rs2941515 is the last nucleotide of the coding exon 3 of *STARD3*. HSF predicted that rs2941515 and rs73296109 would potentially alter an exonic splicing enhancer (ESE) site and activate an exonic cryptic activator site.

Rs8738 near the *SETD4* gene is a synonymous variant enriched 3.10 times. Finally, rs11597050 is a synonymous variant in *SGPL1* enriched 4.53 times. Table 2 shows the enrichment and Odds ratio of effect alleles of these SNPs. Aforementioned rs8738 was also enriched in the Italian cohort of sIBM patients (allelic count: 5/24). HSF predicted that both these synonymous SNPs, rs8738 and rs11597050, would potentially alter splicing.

Two SNPs in *NOTCH4*, rs115393945 (previously reported as rs443198) and rs915894 were found with a frequency of 0.43 and 0.5 in the sIBM cohort as compared to 0.30 and 0.36 respectively in the control population. The association of these SNPs did not reach statistical significance (p-value>0.05).
Discussion

WES data of 30 Finnish sIBM patients did not reveal any putative causative variants in 180 known myopathy genes [24] including \textit{SQSTM1} and \textit{VCP} (more than 99\% bases covered 10x or above). The previously reported SNPs in \textit{SQSTM1} (rs104893941, rs147810437, rs11548633 and rs373585056) and in \textit{VCP} (rs140913250 and rs387906789) [28] were not found in our cohort of 30 sIBM patients. We did not find any other rare variants in these two genes.

Thereafter, we designed a case-control based study to look for association of SNPs with a complex multifactorial phenotype like sIBM. The genetics of the Finnish population is well suited for this kind of approach due to its relative genetic isolation and presence of historical bottleneck effects [29]. In Finland, the current best estimate for the prevalence of sIBM is 70 per million [30]. Our study suggests the association of 7 SNPs with OR>2 and a nominal p-value<0.005, leading to a higher risk of sIBM in Finnish patients. Two SNPs, rs1063318 in \textit{HLA-DQB1} and rs7383287 in \textit{HLA-DOB} enriched in the study population locate in the HLA locus. This is in concordance with the earlier published studies associating HLA haplotypes with increased susceptibility for sIBM [15-18, 31-33].

Five other SNPs are located in the \textit{STARD3}, \textit{SGPL1} and \textit{SETD4} genes. These genes are involved in cellular processes such as methyltransferase activity, sphingolipid transport and metabolism. All these SNPs were predicted to result in broken exonic enhancer sites, potentially altering the splicing.

In the \textit{STARD3} gene rs2941515, rs73296109 and rs2941513 represent a haplotype (about 5 kb). Rs2941515 in particular is the last nucleotide of the coding exon 3 of the gene. \textit{STARD3} belongs to a subfamily of lipid trafficking proteins related to the transport of cholesterol and sphingolipids [34]. Changes in this gene may suggest involvement of intra-cellular trafficking of sphingolipids and sterols in sIBM.
Rs11597050 is located in the *SGPL1* gene. SGPL1 is a protein that cleaves phosphorylated sphingoid bases (PSBs) into fatty aldehydes and phosphoethanolamine as the final enzyme in sphingolipid catabolism [35]. SGPL1 also elevates stress-induced ceramide production and apoptosis. Abnormal sphingolipid alterations have previously been observed in a wide range of muscular phenotypes[36]. Considering their important role abnormal synthesis, metabolism or degradation of sphingolipids, might be a predisposing factor in sIBM as well as a number of disorders yet to be clearly understood [36].

The association of these 4 SNPs found in two different genes (*STARD3* and *SGPL1*) may underline the importance of sphingolipid metabolism and trafficking in sIBM pathology.

Rs8738 in *SETD4* was found to be enriched in the Finnish and the Italian cohort of sIBM patients. *SETD4* is a gene belonging to the putative methyltransferase family and the subfamily of Set domain containing proteins. The protein SETD4 is a cytosolic and nuclear Lysine methyltransferase involved in cell cycle regulation [37].

A possible limitation in this study is represented by the different enrichment methods which were used to capture the exomes of the 30 sIBM patients and 193 controls. This difference created a possibility of batch effects when genotyping all 223 exomes together. To counter this, BAM files were generated only for the shared probe regions, a total of 35,347 SNP markers, thus losing a small number of SNP markers for which the association cannot be ascertained in the present design and this should be taken into account in the subsequent studies. Whether the enriched variants identified could increase susceptibility by themselves or in linkage disequilibrium with other risk variants also warrants further consideration. Increasing the power by including additional cases from Finland, and for comparison other populations, could reveal even more specific regions or genes with increased susceptibility for sIBM.
Despite the challenges of sIBM being a complex disorder, genetic research is likely to result in important findings for understanding the etiology of the disease. Unusual combinations of different mutated genes in the same individual may result in higher risk of developing a multifactorial sporadic disease like sIBM.
Acknowledgments

The authors would like to thank Merja Soininen and Helena Luque, for technical help and Sini Penttilä, Tiina Suominen and Sara Lehtinen for acquisition of samples. Jaakko Sarparanta for his critical revision of the manuscript. This study was supported by the Folkhälsan Research Foundation, Finska Läkaresällskapet and the Jane and Aatos Erkko Foundation, (B.U).
Conflicts of interests

The authors report no conflicts of interests.
References


**Figure Legends**

Figure 1: Enrichment of SNPs in sIBM

Seven enriched SNPs were identified from the case–control association study in the sIBM population. A comparison of the frequency for each SNP in the control population and the sIBM cohort is shown here.
Table 1 List of variants found in WES data of 30 sIBM patients.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Total no of variants</th>
<th>Non-synonymous and non-intronic variants</th>
<th>Variants in Myocap genes* (LOF variants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1378</td>
<td>778</td>
<td>12 (0)</td>
</tr>
<tr>
<td>2</td>
<td>1356</td>
<td>756</td>
<td>12 (0)</td>
</tr>
<tr>
<td>3</td>
<td>1404</td>
<td>802</td>
<td>12 (0)</td>
</tr>
<tr>
<td>4</td>
<td>1317</td>
<td>737</td>
<td>5 (0)</td>
</tr>
<tr>
<td>5</td>
<td>1339</td>
<td>743</td>
<td>5 (0)</td>
</tr>
<tr>
<td>6</td>
<td>1488</td>
<td>845</td>
<td>8 (0)</td>
</tr>
<tr>
<td>7</td>
<td>1385</td>
<td>770</td>
<td>15 (0)</td>
</tr>
<tr>
<td>8</td>
<td>1385</td>
<td>786</td>
<td>9 (1)</td>
</tr>
<tr>
<td>9</td>
<td>1397</td>
<td>763</td>
<td>6 (0)</td>
</tr>
<tr>
<td>10</td>
<td>1443</td>
<td>802</td>
<td>12 (0)</td>
</tr>
<tr>
<td>11</td>
<td>1440</td>
<td>806</td>
<td>7 (0)</td>
</tr>
<tr>
<td>12</td>
<td>1463</td>
<td>807</td>
<td>10 (0)</td>
</tr>
<tr>
<td>13</td>
<td>1486</td>
<td>820</td>
<td>15 (0)</td>
</tr>
<tr>
<td>14</td>
<td>1321</td>
<td>756</td>
<td>8 (0)</td>
</tr>
<tr>
<td>15</td>
<td>1393</td>
<td>786</td>
<td>7 (0)</td>
</tr>
<tr>
<td>16</td>
<td>1421</td>
<td>792</td>
<td>7 (0)</td>
</tr>
<tr>
<td>17</td>
<td>1380</td>
<td>765</td>
<td>9 (0)</td>
</tr>
<tr>
<td>18</td>
<td>1404</td>
<td>788</td>
<td>11 (1)</td>
</tr>
<tr>
<td>19</td>
<td>1455</td>
<td>825</td>
<td>17 (0)</td>
</tr>
<tr>
<td>20</td>
<td>1399</td>
<td>783</td>
<td>9 (0)</td>
</tr>
<tr>
<td>21</td>
<td>1407</td>
<td>805</td>
<td>11 (0)</td>
</tr>
<tr>
<td>22</td>
<td>1455</td>
<td>816</td>
<td>8 (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>23</td>
<td>1428</td>
<td>804</td>
<td>15 (0)</td>
</tr>
<tr>
<td>24</td>
<td>1428</td>
<td>790</td>
<td>5 (0)</td>
</tr>
<tr>
<td>25</td>
<td>1425</td>
<td>829</td>
<td>13 (0)</td>
</tr>
<tr>
<td>26</td>
<td>1359</td>
<td>808</td>
<td>7 (0)</td>
</tr>
<tr>
<td>27</td>
<td>1417</td>
<td>796</td>
<td>8 (0)</td>
</tr>
<tr>
<td>28</td>
<td>1460</td>
<td>811</td>
<td>11 (1)</td>
</tr>
<tr>
<td>29</td>
<td>1397</td>
<td>778</td>
<td>12 (0)</td>
</tr>
<tr>
<td>30</td>
<td>1373</td>
<td>779</td>
<td>11 (0)</td>
</tr>
</tbody>
</table>

§ Variants with MAF less than 1%, read depth more than 20 reads and a reference to allele ratio more than 0.25 were included.

* Ref 24, ** LOF = Loss of function variants.
### Table 2 Effect allele frequency (EAF)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genomic coordinates (hg19)</th>
<th>Nearest gene</th>
<th>Cases</th>
<th>Controls</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1063318</td>
<td>6:32632745</td>
<td><em>HLA-DQB1</em></td>
<td>0.65</td>
<td>0.32</td>
<td>3.89</td>
<td>2.1804 to 6.9383</td>
<td>0.0001</td>
</tr>
<tr>
<td>rs7383287</td>
<td>6:32783086</td>
<td><em>HLA-DOB</em></td>
<td>0.40</td>
<td>0.17</td>
<td>3.22</td>
<td>1.7829 to 5.8025</td>
<td>0.0002</td>
</tr>
<tr>
<td>rs2941515</td>
<td>17:37813338</td>
<td><em>STARD3</em></td>
<td>0.22</td>
<td>0.05</td>
<td>4.83</td>
<td>2.2197 to 10.4886</td>
<td>0.0002</td>
</tr>
<tr>
<td>rs73296109</td>
<td>17:37815749</td>
<td><em>STARD3</em></td>
<td>0.22</td>
<td>0.05</td>
<td>4.83</td>
<td>2.2197 to 10.4886</td>
<td>0.0002</td>
</tr>
<tr>
<td>rs2941513</td>
<td>17:37818561</td>
<td><em>STARD3</em></td>
<td>0.22</td>
<td>0.05</td>
<td>4.83</td>
<td>2.2197 to 10.4886</td>
<td>0.0002</td>
</tr>
<tr>
<td>rs8738</td>
<td>21:37420650</td>
<td><em>SETD4</em></td>
<td>0.23</td>
<td>0.08</td>
<td>3.74</td>
<td>1.8120 to 7.7085</td>
<td>0.0007</td>
</tr>
<tr>
<td>rs11597050</td>
<td>10:72614524</td>
<td><em>SGPL1</em></td>
<td>0.15</td>
<td>0.03</td>
<td>5.15</td>
<td>2.0335 to 13.0413</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

OR= Odds ratio; CI=Confident Interval
* For each SNP a p-value was calculated using Fisher exact test.