


Association study reveals novel risk loci for sporadic inclusion body myositis

M. Johari^a , M. Arumilli^{a,b,c}, J. Palmio^d, M. Savarese^a, G. Tasca^e, M. Mirabella^f, N. Sandholm^{a,g,h}, H. Lohi^{a,b,c}, P. Hackman^a and B. Udd^{a,d,i}

^aFolkhälsan Institute of Genetics, Medicum, University of Helsinki, Helsinki; ^bResearch Programs Unit, Molecular Neurology, University of Helsinki, Helsinki; ^cDepartment of Veterinary Biosciences, University of Helsinki, Helsinki; ^dNeuromuscular Research Center, Tampere University and University Hospital, Tampere, Finland; ^eInstitute of Neurology, Policlinico 'A. Gemelli' Foundation University Hospital, Rome; ^fInstitute of Neurology, Catholic University School of Medicine, Rome, Italy; ^gAbdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki; ^hResearch Program Unit, Diabetes and Obesity, University of Helsinki, Helsinki; and ⁱDepartment of Neurology, Vaasa Central Hospital, Vaasa, Finland

Keywords:

association study,
case–control study,
genetic risk factors,
HLA, risk loci,
sphingolipids, sporadic
inclusion body myositis,
whole exome sequencing

Received 14 November 2016

Accepted 4 January 2017

*European Journal of
Neurology* 2017, **0**: 1–6

doi:10.1111/ene.13244

Background and purpose: The aim was to identify potential genetic risk factors associated with sporadic inclusion body myositis (sIBM).

Methods: An association based case–control approach was utilized on whole exome sequencing data of 30 Finnish sIBM patients and a control cohort ($n = 193$). A separate Italian cohort of sIBM patients ($n = 12$) was used for evaluation of the results.

Results: Seven single nucleotide polymorphisms were identified in five genes that have a considerably higher observed frequency in Finnish sIBM patients compared to the control population, and the previous association of the genetic human leukocyte antigen region was confirmed.

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

Introduction

Sporadic inclusion body myositis (sIBM) is a complex, probably 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 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 cytochrome-oxidase-negative succinate-dehydrogenase-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 major histocompatibility complex (MHC) antigens is part of the immune response [11].

One recent hypothesis suggests that abnormal multi-protein aggregates cause the degeneration of aging

Correspondence: P. Hackman, Department of Medical Genetics, Folkhälsan Institute of Genetics, University of Helsinki, Biomedicum, Haartmaninkatu 8-Pb 63, Helsinki FI-00014, Finland (tel.: +358-294125069; fax: +358-919125073; e-mail: peter.hackman@helsinki.fi).

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 human leukocyte antigens (HLAs). *HLA-DR3* and the extended 8.1 ancestral haplotype have been strongly associated with increased susceptibility for sIBM [14]. 35.2AH and 52.1AH 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 (odds ratio > 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 homogeneous Finnish population background, a whole exome sequencing (WES) based case–control study was designed and an enrichment of certain specific variants was observed 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 European Neuromuscular Centre criteria and the diagnosis was classified as clinico-pathologically defined IBM [7]. Muscle magnetic resonance imaging, 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 a standard procedure.

Genetic and bioinformatic analysis

Whole exome sequencing 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 were downloaded from the 1000 Genomes Project website (<http://www.internationalgenome.org/>). To counter 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 Burrows-Wheeler Aligner (BWA) [22]. Duplicate reads were marked using Picard (<https://broadinstitute.github.io/picard>) and local realignment was performed using Genome Analysis Toolkit (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]. All the common variants [minor allele frequency (MAF) higher than 1%] in the 1000 Genomes and ExAC database and our internal control cohort were filtered out.

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 $P > 0.000001$ 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. The 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]. The 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.

The 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 seven significant identified single nucleotide polymorphisms (SNPs). All significantly enriched SNPs were confirmed by Sanger sequencing.

Results

The WES data of the 30 Finnish sIBM patients were analyzed 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 seven SNPs were identified with a nominal P value <0.005 enriched in the Finnish sIBM population (Fig. 1).

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

Table 1 List of variants found in whole exome sequencing data of 30 sporadic inclusion body myositis patients

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

LOF, loss of function. ^aVariants with minor allele frequency less than 1%, read depth more than 20 reads and a reference to allele ratio more than 0.25 were included; ^bReference 24.

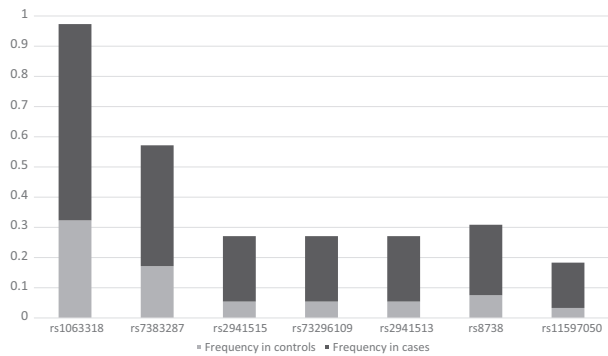


Figure 1 Enrichment of SNPs in sIBM. Seven enriched SNPs were identified from the case–control association study in the sIBM population. A comparison of the frequency for each SNP in the control population and the sIBM cohort is shown here.

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 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. The 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 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

Whole exome sequencing data of 30 Finnish sIBM patients did not reveal any putative causative variants in 180 known myopathy genes [24] including *SQSTM1* and *VCP* (more than 99% bases covered $10\times$ or above). The previously reported SNPs in *SQSTM1* (rs104893941, rs147810437, rs11548633 and rs373585056) and in *VCP* (rs140913250 and rs387906789) [28] were not found in our cohort of 30 sIBM patients. No other rare variants were found in these two genes.

Table 2 Effect allele frequency

SNP	Genomic coordinates (hg19)	Nearest gene	Cases	Controls	OR	95% CI	<i>P</i> value ^a
rs1063318	6:32 632 745	<i>HLA-DQB1</i>	0.65	0.32	3.89	2.1804–6.9383	0.0001
rs7383287	6:32 783 086	<i>HLA-DOB</i>	0.40	0.17	3.22	1.7829–5.8025	0.0002
rs2941515	17:37 813 338	<i>STARD3</i>	0.22	0.05	4.83	2.2197–10.4886	0.0002
rs73296109	17:37 815 749	<i>STARD3</i>	0.22	0.05	4.83	2.2197–10.4886	0.0002
rs2941513	17:37 818 561	<i>STARD3</i>	0.22	0.05	4.83	2.2197–10.4886	0.0002
rs8738	21:37 420 650	<i>SETD4</i>	0.23	0.08	3.74	1.8120–7.7085	0.0007
rs11597050	10:72 614 524	<i>SGPL1</i>	0.15	0.03	5.15	2.0335–13.0413	0.0011

SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval. ^aFor each SNP a *P* value was calculated using the Fisher exact test.

Thereafter, a case–control based study was designed to look for the 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 the 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 seven SNPs with odds ratio >2 and a nominal *P* value <0.005, leading to a higher risk of sIBM in Finnish patients. Two SNPs, rs1063318 in *HLA-DQB1* and rs7383287 in *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 *STARD3*, *SGPL1* and *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 *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. *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 intracellular trafficking of sphingolipids and sterols in sIBM.

Rs11597050 is located in the *SGPL1* gene. *SGPL1* is a protein that cleaves phosphorylated sphingoid bases 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 four 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 (Supplementary data S1), 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 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 a higher risk of developing a multifactorial sporadic disease like sIBM.

Acknowledgements

The authors would like to thank Merja Soininen and Helena Luque for technical help, Sini Penttilä, Tiina

Suominen and Sara Lehtinen for acquisition of samples, and 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 Erkkö Foundation.

Disclosure of conflicts of interests

The authors declare no financial or other conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Bed file listing common probe regions.

References

1. Needham M, Mastaglia FL. Sporadic inclusion body myositis: a review of recent clinical advances and current approaches to diagnosis and treatment. *Clin Neurophysiol* 2016; **127**: 1764–1773.
2. Needham M, Corbett A, Day T, Christiansen F, Fabian V, Mastaglia FL. Prevalence of sporadic inclusion body myositis and factors contributing to delayed diagnosis. *J Clin Neurosci* 2008; **15**: 1350–1353.
3. Cox FM, Reijnierse M, van Rijswijk CS, Wintzen AR, Verschuuren JJ, Badrising UA. Magnetic resonance imaging of skeletal muscles in sporadic inclusion body myositis. *Rheumatology (Oxford)* 2011; **50**: 1153–1161.
4. Cox FM, Verschuuren JJ, Verbist BM, Niks EH, Wintzen AR, Badrising UA. Detecting dysphagia in inclusion body myositis. *J Neurol* 2009; **256**: 2009–2013.
5. Dimachkie MM, Barohn RJ. Inclusion body myositis. *Curr Neurol Neurosci Rep* 2013; **13**: 321.
6. Lloyd TE, Mammen AL, Amato AA, Weiss MD, Needham M, Greenberg SA. Evaluation and construction of diagnostic criteria for inclusion body myositis. *Neurology* 2014; **83**: 426–433.
7. Rose MR, Group EIW. 188th ENMC International Workshop: inclusion body myositis, 2–4 December 2011, Naarden, The Netherlands. *Neuromuscul Disord* 2013; **23**: 1044–1055.
8. Tasca G, Monforte M, De Fino C, Kley RA, Ricci E, Mirabella M. Magnetic resonance imaging pattern recognition in sporadic inclusion-body myositis. *Muscle Nerve* 2015; **52**: 956–962.
9. Askanas V, Engel WK. Inclusion-body myositis: newest concepts of pathogenesis and relation to aging and Alzheimer disease. *J Neuropathol Exp Neurol* 2001; **60**: 1–14.
10. Oldfors A, Moslemi AR, Jonasson L, Ohlsson M, Kollberg G, Lindberg C. Mitochondrial abnormalities in inclusion-body myositis. *Neurology* 2006; **66**(Suppl.1): S49–S55.
11. Muntzing K, Lindberg C, Moslemi AR, Oldfors A. Inclusion body myositis: clonal expansions of muscle-infiltrating T cells persist over time. *Scand J Immunol* 2003; **58**: 195–200.
12. Askanas V, Engel WK. Molecular pathology and pathogenesis of inclusion-body myositis. *Microsc Res Tech* 2005; **67**: 114–120.
13. Askanas V, Engel WK, Nogalska A. Sporadic inclusion-body myositis: a degenerative muscle disease associated with aging, impaired muscle protein homeostasis and abnormal mitophagy. *Biochim Biophys Acta* 2015; **1852**: 633–643.
14. Garlepp MJ, Laing B, Zilko PJ, Ollier W, Mastaglia FL. HLA associations with inclusion body myositis. *Clin Exp Immunol* 1994; **98**: 40–45.
15. Koffman BM, Sivakumar K, Simonis T, Stroncek D, Dalakas MC. HLA allele distribution distinguishes sporadic inclusion body myositis from hereditary inclusion body myopathies. *J Neuroimmunol* 1998; **84**: 139–142.
16. Mastaglia FL, Needham M, Scott A, et al. Sporadic inclusion body myositis: HLA-DRB1 allele interactions influence disease risk and clinical phenotype. *Neuromuscul Disord* 2009; **19**: 763–765.
17. Scott AP, Allcock RJ, Mastaglia F, Nishino I, Nonaka I, Laing N. Sporadic inclusion body myositis in Japanese is associated with the MHC ancestral haplotype 52.1. *Neuromuscul Disord* 2006; **16**: 311–315.
18. Rojana-udomsart A, Mitrpant C, James I, et al. Analysis of HLA-DRB3 alleles and supertypal genotypes in the MHC Class II region in sporadic inclusion body myositis. *J Neuroimmunol* 2013; **254**: 174–177.
19. Scott AP, Laing NG, Mastaglia F, Dalakas M, Needham M, Allcock RJ. Investigation of NOTCH4 coding region polymorphisms in sporadic inclusion body myositis. *J Neuroimmunol* 2012; **250**: 66–70.
20. Gang Q, Bettencourt C, Machado PM, et al. The effects of an intronic polymorphism in TOMM40 and APOE genotypes in sporadic inclusion body myositis. *Neurobiol Aging* 2015; **36**: 1766.e1–e3.
21. Wehl CC, Baloh RH, Lee Y, et al. Targeted sequencing and identification of genetic variants in sporadic inclusion body myositis. *Neuromuscul Disord* 2015; **25**: 289–296.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
23. Erlich RL, Jia X, Anderson S, et al. Next-generation sequencing for HLA typing of class I loci. *BMC Genom* 2011; **12**: 42.
24. Evilä A, Arumilli M, Udd B, Hackman P. Targeted next-generation sequencing assay for detection of mutations in primary myopathies. *Neuromuscul Disord* 2016; **26**: 7–15.
25. Karssen LC, van Duijn CM, Aulchenko YS. The GenABEL Project for statistical genomics. *F1000Research* 2016; **5**: 514.
26. Klionsky DJ, Abdelmohsen K, Abe A, et al. Guidelines for the use and interpretation of assays for monitoring autophagy (3rd edition). *Autophagy* 2016; **12**: 1–222.
27. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.
28. Gang Q, Bettencourt C, Machado PM, et al. Rare variants in SQSTM1 and VCP genes and risk of sporadic inclusion body myositis. *Neurobiol Aging* 2016; **47**: 218.e1–218.e9.
29. Peltonen L, Jalanko A, Varilo T. Molecular genetics the Finnish disease heritage. *Hum Mol Genet* 1999; **8**: 1913–1923.

30. Udd B. Lihastautien diagnostiikka tarkentuu: potilaita Suomessa yli 10000. *Suomen lääkärilehti* 2007; **62**: 519–522.
31. Badrising UA, Schreuder GM, Giphart MJ, *et al.* Associations with autoimmune disorders and HLA class I and II antigens in inclusion body myositis. *Neurology* 2004; **63**: 2396–2398.
32. Price P, Santoso L, Mastaglia F, *et al.* Two major histocompatibility complex haplotypes influence susceptibility to sporadic inclusion body myositis: critical evaluation of an association with HLA-DR3. *Tissue Antigens* 2004; **64**: 575–580.
33. Burnell AM, Houthoofd K, O'Hanlon K, Vanfleteren JR. Alternate metabolism during the dauer stage of the nematode *Caenorhabditis elegans*. *Exp Gerontol* 2005; **40**: 850–856.
34. Tsujishita Y, Hurley JH. Structure and lipid transport mechanism of a StAR-related domain. *Nat Struct Mol Biol* 2000; **7**: 408–414.
35. Reina E, Camacho L, Casas J, Van Veldhoven PP, Fabrias G. Determination of sphingosine-1-phosphate lyase activity by gas chromatography coupled to electron impact mass spectrometry. *Chem Phys Lipids* 2012; **165**: 225–231.
36. Lamari F, Mochel F, Sedel F, Saudubray JM. Disorders of phospholipids, sphingolipids and fatty acids biosynthesis: toward a new category of inherited metabolic diseases. *J Inherit Metab Dis* 2013; **36**: 411–425.
37. Faria JA, Correa NC, de Andrade C, *et al.* SET domain-containing protein 4 (SETD4) is a newly identified cytosolic and nuclear lysine methyltransferase involved in breast cancer cell proliferation. *J Cancer Sci Ther* 2013; **5**: 58–65.