In silico analysis and 3-D structure prediction of non-synonymous single nucleotide polymorphisms (nsSNPs) in the human *MED12* gene associated with uterine leiomyomas

ABSTRACT

Aims. Uterine leiomyomas are one of the most common benign gynecologic tumors, but the exact causes are not completely understood. *MED12* mutation was discovered in approximately 71% of uterine leiomyomas. Our recent studies confirmed the high frequency of *MED12* exon2 deleterious mutations in uterine leiomyoma in senegalese patients. In this context,molecular dynamics simulations of wild-type and non synonymous variants were conducted to access their structural dynamic and stability characteristics.

Methodology. Uterine leiomyoma tissues were obtained from symptomatic women who underwent hysterectomy or myomectomy for medically indicated reasons. We collected 50 uterine leiomyomas and sequenced *MED12* exon2 region after DNA extration and amplification. To understand the mutation's role in this gene, we utilized computational tools based on different algorithms, including missense tools to predict pathogenic variants, stability tools to analyze the impression of a mutated gene on the function of the protein, and molecular architecture research analysis to detect the gene intearctions.

Results. A significant genetic alteration of the *MED12* gene has showed with a high frequency of mutations noted in particular codon 44 of exon 2. All these mutations being predicted as deleterious testify to their implication in the pathobiology of uterine fibroids. In addition, the noted alterations lead to instability of the MED12 protein and thus a change in its biological function in uterine fibroids.

Key words: MED12 ; leiomyomas ; *in silico* prediction ; 3D structure modeling

INTRODUCTION

Uterine fibroids, more commonly known as myomas, arethe most common benign tumours of female reproductiveorgans [1]. They are associated with significant morbidity andconstitute a real public health problem. Several studies have also indicated theinvolvement of hormonal status in the development of uterinefibroids [2,3]. Although themolecular mechanisms involved in their etiology remainincompletely elucidated, several studies have focused on the*MED12* gene [4]. Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator functions as a bridge to convey information from gene-specific regulatory proteins to the basal RNA polymerase II transcription machinery. Mediator is recruited to promoters by direct interactions with regulatory proteins and serves as a scaffold for the assembly of a functional preinitiation complex with RNA polymerase II and the general transcription factors. This subunit may specifically regulate transcription of targets of the Wnt signaling pathway and SHH signaling pathway [5,6].

One the major challenges in modern genetics is predicting the effect of the overwhelming number of variants being revealed through sequencing projects. This is particularly important in analyzing variants occurring in the human population that could be involved in the pathogenesis of disease. Since the structure of a protein is intimately linked to its stability, function and interactions, many *in silico* prediction methods employ knowledge of protein structure, either exclusively or in combination with sequence-based features, with the aim of providing high-quality predictions [7,8,9]. Knowledge of protein structure can be used to predict the phenotypic consequence of a missense variant. It has been noted [3] that in many studies the challenges and costs arise more from the analysis of the data than the actual sequencing. The aims were to provide an analysis describing structurally damaging changes introduced by human missense variants, which can suggest that a variant is likely to be disease causing, and to compare results obtained using experimental and 3D model structures.

1. METHODS

1.1. Samples

50 patients with uterine fibroids were recruited from the Grand Yoff General Hospital. After obtaining informed consent, each patient was given a biopsy of the tumour tissue and a blood sample (to serve as a control) on an EDTA tube.

1.2. DNA Extraction, *MED12* Gene Amplification and Sequencing

The total DNA of each sample was extracted using the Qiagen protocol (Qiagen Dneasy Blood and Tissues kit).After extraction, exon 2 of the *MED12* gene wasamplified. Primers forward 5'GCCCTTTCACCTTGTTCCTT3' and reverse 5'TGTCCCTATAAGTCTTCCCAACC3' were used for PCR reaction under theconditions previously described by Mäkinen et al., [4].

Sequencingreactions were performed in a MJ Research PTC-225 Peltier thermocycler with ABIPRISM BigDye TM Terminator Cycle kits. Each sample was sequenced using forward primer.

1.3. Detection of Mutations

To determine the presence of any mutation and its position relative to the *MED12* gene, the raw sequencing data were submitted to the Mutation Surveyor software version 5.0.1(www.softgenetics.com), which compares the submitted chromatograms with thereference sequence of *MED12* gene (NT_011669_70337906). Non-synonymous deleterious mutations were analysis as previously described by Keneme et al.,[10].

1.4. Prediction of deleterious nsSNP

The functional impact of variants may lead to a wide range of molecular changes, even within a single protein, including disrupted stability and structure, disrupted macromolecular binding, ablation of post-translational modification (PTM) sites, among others. To help prioritize potentially deleterious variants, groups have developed *in silico* prediction programs. Many of these methods rely primarily or exclusively on sequence-based features, such as amino-acid evolutionary conservation.

- **ClinVar** (http://www.ncbi.nlm.nih.gov/clinvar/) provides a freely available archive of reports of relationships among medically important variants and phenotypes [11]. ClinVar accessions submissions reporting human variation, interpretations of the relationship of that variation to

human health and the evidence supporting each interpretation. For this analysis the reference NP_005111.2 was used to find clinical significance of nsSNP.

- **ProtParam** computes various physico-chemical properties that can be deduced from a protein sequence [12]. The protein can either be specified as a Swiss-Prot/TrEMBL accession number or ID, or in form of a raw sequence. Using this tool it is possible to compute an instability index. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.
- **Missense 3D** can be used to predict the phenotypic consequence of a missense variant [13]. The Mutant and Wild-type structures were analyzed to identify whether the structural consequence of the substitution is expected to be damaging in terms of the stability of the folded protein. Based on well-established principles of protein conformation and previous studies on the structural consequences of disease-associated substitutions, 17 structural features were considered. The aims were to provide an analysis describing structurally damaging changes introduced by human missense variants, which can suggest that a variant is likely to be disease causing, and to compare results obtained using experimental and 3D model structures.
- **MutPred2** is a machine learning-based method, and software package that integrates genetic and molecular data to reason probabilistically about the pathogenicity of amino acid substitutions [14]. This is achieved by providing (1) a general pathogenicity prediction, and (2) a ranked list of specific molecular alterations potentially affecting the phenotype. The loss and gain of structural and functional properties are modelled via posterior probabilities. Through assessment of these probabilities, MutPred can predict the molecular cause of disease-associated substitution. Analysis was done with wild-type protein sequence in FASTA format ID NP_00511.2 and the substitution sites were identified. The probability of the mutation being deleterious is reported. The final prediction score is the average of the scores from all networks and ranges between 0 and 1; higher scores reflect a higher probability of pathogenicity. Any molecular mechanisms that are likely to be disrupted due to the mutation are reported, with corresponding P value.Functional analysis includes the prediction of DNA-binding site, catalytic domains, calmodulin-binding targets, and post-translational modification sites [15-17].

1.5. Prediction effects of nsSNP on protein stability

To see whether the nsSNP of exon 2 may affect the stability of MED12 protein, each mutant was compared to wild type using *in silico* prediction softwares. The Uniprot ID (Q93074 or MED12_HUMAN) and AlphaFold structure prediction for PDB file (AF-Q93074-F1) were used for analysis.

- **DynaMut2** is a web server that combines Normal Mode Analysis (NMA) methods to capture protein motion and graph-based signatures to represent the wild-type environment to investigate the effects of single and multiple point mutations on protein stability and dynamics [18,19]. DynaMut2 was able to accurately predict the effects of missense mutations on protein stability, achieving Pearson's correlation of up to 0.72 (RMSE: 1.02 kcal/mol) on a single point

and 0.64 (RMSE: 1.80 kcal/mol) on multiple-point missense mutations across 10-fold cross-validation and independent blind tests. For single-point mutations, DynaMut2 achieved comparable performance with other methods when predicting variations in Gibbs Free Energy (ΔΔG) and in melting temperature (ΔTm) [20]. Value binding free energy change (ΔΔG) for mutation were calculated in kcal/mol. Calculation of correlation matrix is frequently utilized to illustrate dynamical information of proteins in two dimension. To observe the correlation in the dynamics, correlation matrices for each of the nsSNP were calculated through DynaMut web server.

- **NetsurfP3**, after the analysis of protein modifications, the structural analysis of protein was predicted using the tool NetSurf-3.0 (http://www.cbs.dtu.dk/services/NetSurfP/) [21]. This tool works on a neural network algorithm and predicts the secondary structure of amino acids in a sequence, as well as its structural disorder and backbone dihedral angles for each residue in the sequence. This server predicts the surface accessibility with Relative Surface Area (RSA), Accessible Surface Area (ASA, secondary structure, disorder, and phi / psi dihedral angles of amino acids in an amino acid sequence. Amino acid sequence of MED12_Human was retrieved with its genbank accession no NP_005111.2 and a length of 2177 amino acids. Each nsSNP was inserted manually to assess its impact on protein function.

2. RESULTS

2.1 *MED12* exon 2 mutations

Out of 50 patients with uterine fibroids, 88% have mutations in exon2 of the *MED12* gene. All single-position mutations affecting exon 2 induce an amino acid change. The frequency of non-synonymous mutations identified are listed in Table 1.

Interpretations of variants in ClinVar show that all nsSNP exceptF45V, K60M and N61Y are associated with uterine leiomyomas (Table 2).

Table 1. *MED12* mutation frequencies

Mutation position	Amino acid change	Number of patients	Frequence
c.107T>G	p.L36R	1	2%
c.128A>C	p.Q43P	1	2%
c.130G>A	p.G44S	12	24%
c.130G>T	p.G44C	4	8%
c.130G>C	p.G44R	3	6%
c.131G>A	p.G44D	10	20%
c.131G>T	p.G44V	3	6%
c.131G>C	p.G44A	1	2%
c.133T>G	p.F45P	1	2%
c.179A>T	p.K60M	1	2%
c.181A>T	p.N61Y	1	2%

Table 2. Clinical significance of nsSNP

Protein	Molecular consequence	Clinical significance
p.L36R	Missense	Associated with leiomyomas (SCV000109663.1)
p.Q43P	Missense	Associated with leiomyomas (SCV000109676.1)
p.G44S	Missense	Others (SCV000599938.1)
p.G44C	Missense	Associated with leiomyomas (SCV000109681.1)
p.G44R	Missense	Associated with leiomyomas (SCV000109680.1)

p.G44D	Missense	WILMS Tumor (SCV000599933.1)
p.G44V	Missense	Associated with leiomyomas (SCV000109684.1)
p.G44A	Missense	Associated with leiomyomas (SCV000109683.1)
p.F45P	Missense	Uncertain significance
p.K60M	Missense	Uncertain significance
p.N61Y	Missense	Uncertain significance

1.2. Effect of nsSNP on protein stability

For each nsSNP, Protparam results identifies the protein with more number of negatively charged amino acids than positive (Table 3). The physio-chemical annotation of the protein revealed the protein as unstable (instability index > 40).

Table 3. Physico-chemical properties of nsSNP using Protparam

Protein	Negatively charged residues	Positively charged residues	Instability index
p.L36R	228	218	57.72
p.Q43P	228	217	57.58
p.G44S	228	217	57.93
p.G44C	228	217	57.63
p.G44R	228	218	57.72
p.G44D	229	217	57.78
p.G44V	228	217	57.69
p.G44A	228	217	57.72
p.F45P	228	217	57.72
p.K60M	228	216	57.72
p.N61Y	228	217	57.72

Analysis using Missense 3D show that no structural damage was detected for nsSNP L36P, F45V, K60M and N61Y. It was identified that all residues in codon 44 were identified to fall in disallowed regions (Table 4).

Table 4. Structural damage detected using Missense 3D

Protein	Structural damage detected
p.L36R	No structural damage detected
p.Q43P	Clash (the mutant structure has a MolProbity clash score higher than 30 compared to the wild-type
p.G44S	Disallowed (the substitution triggers disallowed : the mutant residue is in outlier region while the wild-type residue is in the favoured region)
p.G44C	Disallowed
p.G44R	Disallowed
p.G44D	Disallowed
p.G44V	Disallowed
p.G44A	Disallowed
p.F45P	No structural damage detected
p.K60M	No structural damage detected
p.N61Y	No structural damage detected

The loss and gain of structural and functional properties are modelled via posterior probabilities. Through assessment of these probabilities, MutPred predict the molecular cause of disease-associated substitution. Analysis show that all nsSNP except G44D, K60M and N61Y are associated with molecular mecanism such as gain of intrincic disorder, altered disorder interface, gain or loss of helix, gain or loss of loop and post-translationnal modifications such as gain or loss of methylation, loss of ubiquitylation (Table 5).

Table 5. Pathogenicity prediction and molecular mecanism association of nsSNP using MutPred2

Substitution	MutPred2 score	Affected Prosite and ELM motifs	Molecular mecanism with p ≤ 0.05
p.L36R	0.812	ELME 00146/00147	- Gain of intrinsic disorder (p=0.02) - Altered disorder interface (p=0.04) - Gain of β-factor (p=0.03) - Loss of ubiquitylation at K132 (p=0.03) - Loss of methylation at K32
p.Q43P	0.550	None	- Loss of helix (p=0.04) - Gain of loop (p=0.04) - Gain of methylation at K42 (p=0.04)
p.G44S	0.700	None	- Altered disordered interface (p=0.04)
p.G44C	0.859	None	- Loss of intrinsic disorder (p=0.03) - Altered disordered interface (p=0.01) - Gain of helix (p=0.05) - Loss of loop (p=0.02) - Gain of disulfide linkage at G44 (p=0.04)
p.G44R	0.787	None	- Gain of helix (p=0.02) - Altered disordered interface (p=0.05)
p.G44D	0.833	None	
p.G44V	0.847	ELME 000146	- Loss of intrinsic disorder (p=0.04) - Loss of loop (p=0.03)
p.G44A	0.736	ELME 000146	- Altered disordered interface (p=0.03) - Gain of helix (p=0.03) - Loss of loop (p=0.05)
p.F45V	0.619	None	- Altered disordered interface (p=6,7^{e-3}) - Gain of methylation at K42 (p=0.04)
p.K60M	0.339	None	
p.N61Y	0.397	None	

For all nsSNP analysed, only mutations in codon 44 are considered distabilising using Dynamut2 (Table 6). In comparison with the reference sequence, all the mutations show a conformational change in the 3D structure of the MED12 protein; in other words, there is a deformation of the protein and to a large extent a modification of the biological activity of the MED12 protein in uterine fibroids (Figure 1).

Table 6. Protein stability prediction with DynaMut2

Wild-type	Position	Mutant	Chain	Free energy change (ΔΔG	Prediction
L	36	R	A	0.76 Kcal/mol	Stabilising
Q	43	9	A	0.29 Kcal/mol	Stabilising
G	44	S	A	-0.18 Kcal/mol	Distabilising
G	44	C	A	-0.77 Kcal/mol	Distabilising
G	44	R	A	-0.25 Kcal/mol	Distabilising
G	44	D	A	-0.23 Kcal/mol	Distabilising
G	44	V	A	-1.25 Kcal/mol	Distabilising
G	44	A	A	-0.16 Kcal/mol	Distabilising
F	45	V	A	-0.05 Kcal/mol	Distabilising
K	60	M	A	0.21 Kcal/mol	Stabilising
N	61	Y	A	0.29 Kcal/mol	Stabilising

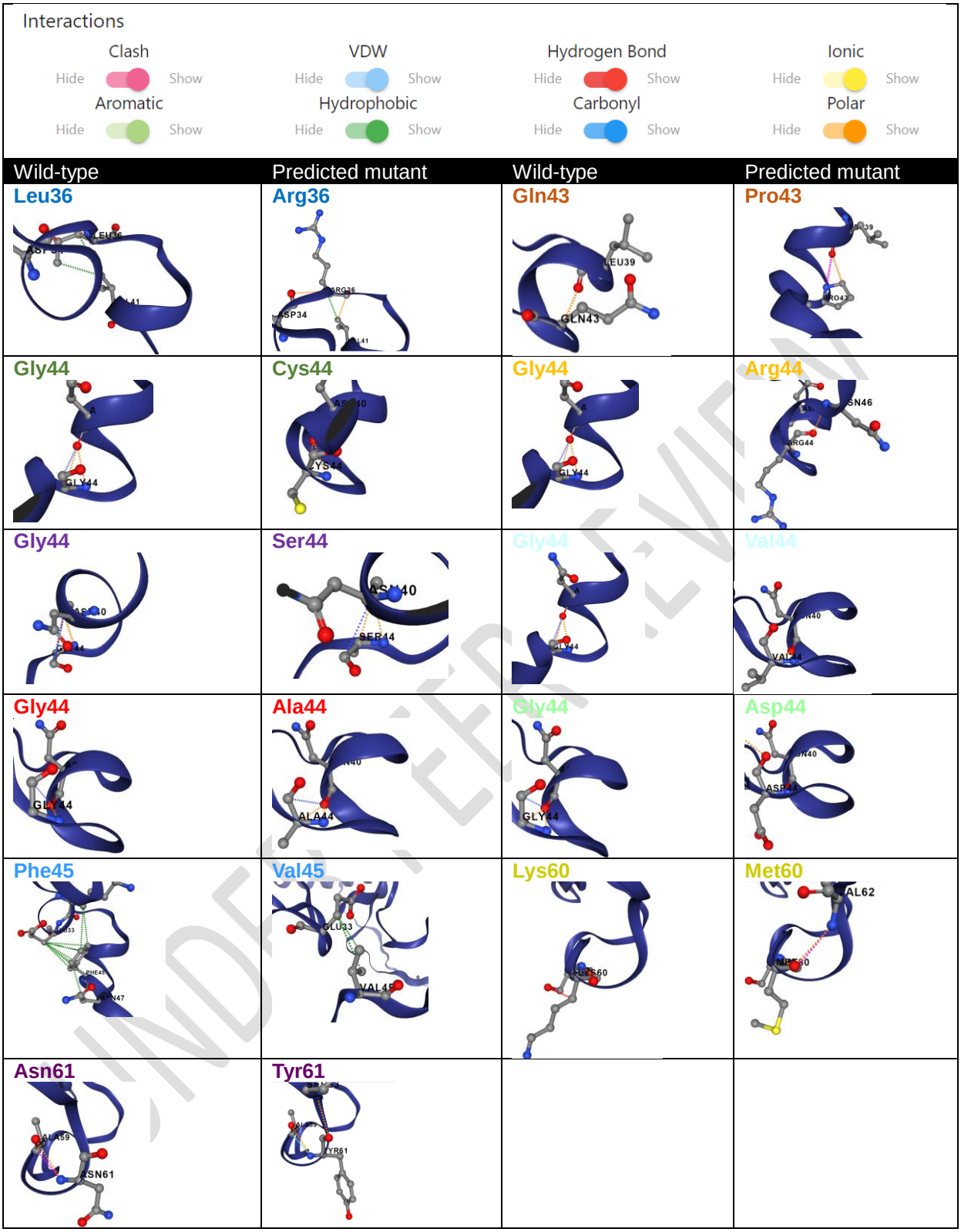

Figure 1. Impact of nsSNP of exon 2 MED12 on 3D-structure of the protein

Normal Mode Analysis (NMA) has been successfully applied to the study of the effects of mutations on protein dynamics (Figure 2).

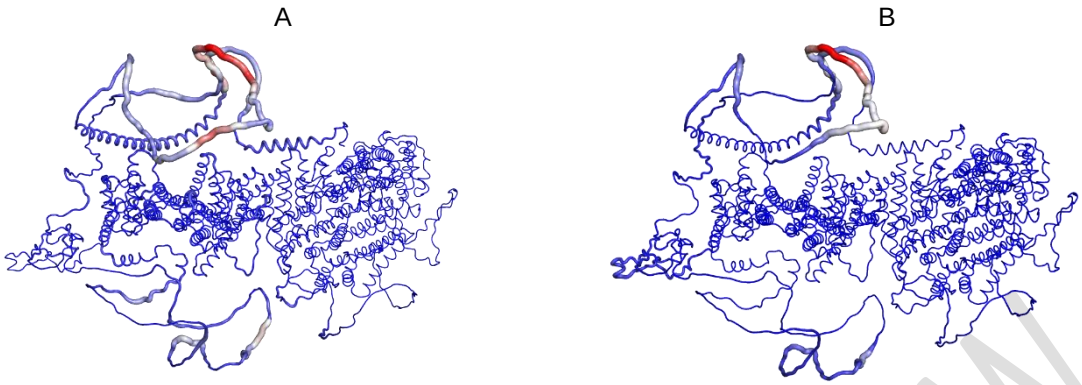

Figure 2. Structural deformation (A) and anatomic fluctuation (B) using Dynamut. The magnitute of the deformation/fluctuation is represented by thin to thick tube colored **blue** (low), white (moderate) and **red** (high).

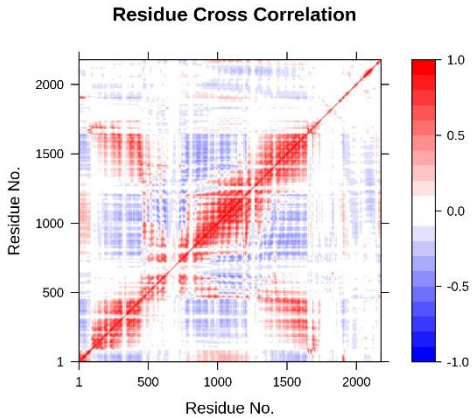

Figure 3. Dynamical Cross-Correlation Map (DCCM)All modes were used to calculated the residue cross-correlation.
Correlation map revealing correlated (**red**) and anti-correlated (**blue**) regions in the protein structure.

NetSurfP tool revealed that surface accessibility of these nsSNPs in secondary structure remains affected due to amino acid substitutions (Table 7). Points mutations are in random coil and exposed regions (Figure 4). Below the secondary structure prediction disorder shows the probability of disorder related to that residue and codon 44 is most affected.

Table 7. Structural analysis of protein prediction using the tool NetSurf-3.0

Protein	**RSA**	**ASA**	**SS**	**P. disorder**
p.L36R	39%	90Å	Coil	68%
p.Q43P	55%	24Å	Coil	67%
p.G44S	23%	27Å	Coil	74%

p.G44C	17%	24Å	Coil	62%
p.G44R	42%	95Å	Coil	83%
p.G44D	40%	58Å	Coil	82%
p.G44V	28%	44Å	Coil	84%
p.G44A	17%	18Å	Coil	68%
p.F45V	32%	49Å	Coil	73%
p.K60M	33%	67Å	Coil	32%
p.N61Y	55%	111Å	Coil	17%

Relative Surface Area (RSA), Accessible Surface Area (ASA), Secondary structure (SS), protein disorder (P disorder)

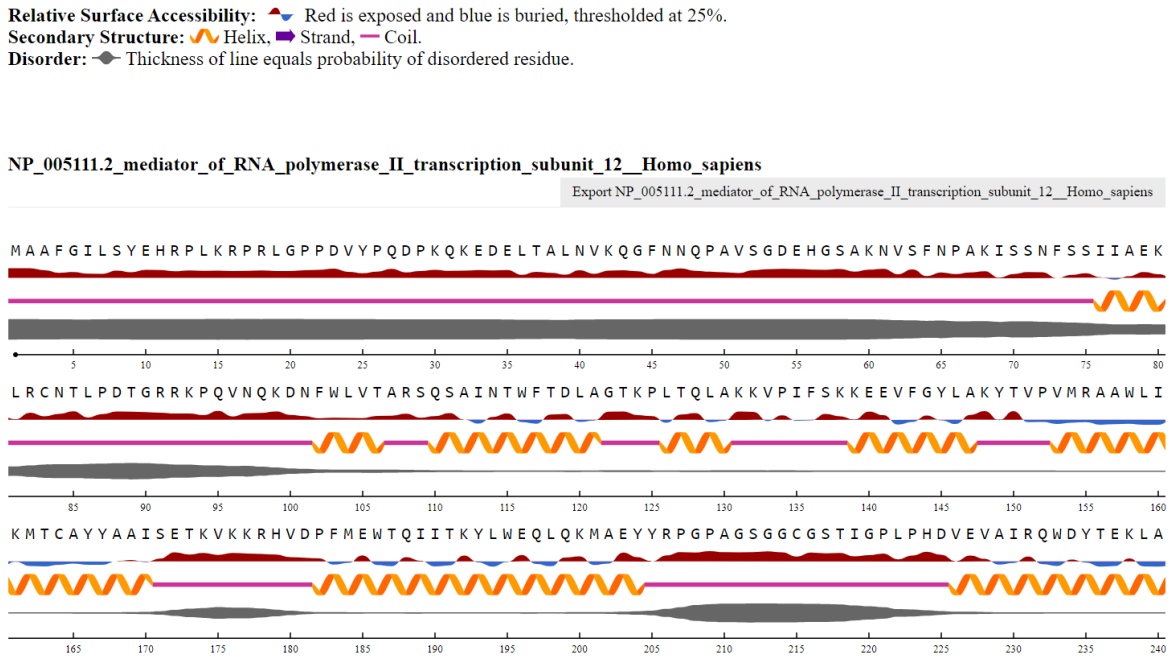

Figure 4. MED12 structural analysis through NetSurf-3.0. Surface Accessibility: The red upward elevation indicates the exposed residue, while the sky blue low elevation shows the buried residue in protein structure.

DISCUSSION

MED12, a geneof 45 exons, located on the X chromosome (Xq13),expressed in all Eukaryotes, forms with Cyclin C, MED13and CDK8, a subunit of this multiprotein complex. In thisstudy, the *MED12* gene has been investigated in Senegalesewomen with uterine fibroids.

Analysis of the chromatogramsrevealed a presence of mutations of the *MED12* gene only intumour tissues with a frequency of 74% (37/50). Thisconfirms the hypothesis that the *MED12* gene is involved inthe occurrence of uterine fibroids. Mäkinen et al. [4] firstdescribed the link between *MED12* mutations and fibroids. In2011, and according to him, mutations in the *MED12* generepresent the largest genetic defect in uterine fibroids. Inaddition, mutations of the *MED12* gene in tumours other thanuterine fibroids are rare. Only 0.3 to 0.5% of colorectalcancers have mutations in the *MED12*

gene, stating that theyare only passenger's mutations [22,23]. 5% of prostatecancers have different MED12 mutations [24,25].

All mutations detected are nsSNP and many of them are associated with uterine leiomyomas by Clinvar. The discovery of pathogenic variants, i.e., variants capable of causing disease, generally relies on a combination of family and population-based sequencing efforts. To assist genetic studies, particularly in characterizing rare variants and dissecting complex disease, machine learning methods have recently been developed to identify the signatures of pathogenicity and to predict the impact of variants of unknown significance. In this contex, changes inthe biological function of MED12 in uterine fibroids are alsohighlighted in protein function prediction analysis. Indeed,the mutations of exon 2 seem to induce gains and/or loss offunction of the MED12 protein. These modificationsconstitute a proof of the biological modifications of the MED12 protein in women with uterine fibroids and thereforetheir implication in the occurrence and/or progression ofthese tumour cells.

All the mutations affecting exon 2appear as deleterious mutations, in particular thoseaffecting codon 44. In other words, all themutations affecting exon 2 cause an aberrant function of the MED12 protein. A studyby Bourbon et al. [26], involving 39 different species,showed that codon 44 is the most conserved codon of the*MED12* gene, which states that this codon plays an importantrole in the biological process of the protein. It has been evidenced that conserved regions are biologically veryimportant, so variations in these regions may lead to potentialfunctional changes.Themissense mutations observed on this codon 44 can render the translated protein non-functional, indicatingthe specific importance of this amino acid for the MED12function with respect to leiomyoma and suggesting that thesemutations could represent alleles gain or loss of function. In Eukaryotes, the Mediator Complex consists of at least 30 proteins [26], structurally divided into four modules, which are the head, the middle, the tail and the kinase modules. The head and middle modules interact directly with RNA polymerase II while the tail module associates with several cofactors to facilitate transcription. The Kinase module interacts with the Mediator Complex to suppress transcription [27]. Indeed, the Mediator Complexexistsin 2 forms. The L. mediatorformcontains 4 modules of the kinase subunit (MED12, MED13, Cyclin C, CDK8 orCDK19) and acts as a receptor. The S. mediator form (without the CDK8 module) stimulates basal transcription. The MED12 domain plays a vital role in connecting Cyclin C-CDK8 to the core of the complex, which activates CDK8 kinase. Moreover,according to the work in reference [27], the binding domainof Cyclin C resides at the level of the N-terminal regionencoded by exons 1 and 2 of the *MED12* gene and codon 44would play a role in this membership. This further confirmsthe transcriptional activation of MED12 aberrant function inuterine fibroids and that codon 44 is essential for this process.

Study the dynamic nature and the role of flexibility/rigidity and accessible conformational landscapes in proteins is essential for understanding their function, as well as to evaluate how changes in a protein might impact its structure, function and interactions, giving rise to different phenotypes. Our analysis of structural impact of nsSNP show that mutations are located at the random coil. Some regions of the

protein chain do not form regular secondary structure and are not characterized by any regular hydrogen bonding pattern. These regions are known as random coils and are found in two locations in proteins: (i) Terminal arms - both at the N-terminus and the C-terminus of the protein; (ii) Loops - Loops are unstructured regions found between regular secondary structure elements [28].Most loops are exposed to the solvent and are have polar or charged side-chains. In some cases loops have a functional role, but in many cases they do not. As a result, loop regions are often poorly conserved (i.e. more prone to change) during evolution.

Additionally, the human proteome is dynamic and changes in response to a stimuli, and post-translational modifications are commonly employed to regulate cellular activity. PTMs occur at distinct amino acid side chains or peptide linkages, and they are most often mediated by enzymatic activity [29,30]. Indeed, it is estimated that 5% of the proteome comprises enzymes that perform more than 200 types of post-translational modifications. These enzymes include kinases, phosphatases, transferases and ligases, which add or remove functional groups, proteins, lipids or sugars to or from amino acid side chains; and proteases, which cleave peptide bonds to remove specific sequences or regulatory subunits. Many proteins can also modify themselves using autocatalytic domains, such as autokinase and autoprotolytic domains.Significant gain or loss of methylation had been observed in L36R, Q43P, F45V variations, and gain of disulfide linkage at G44 in G44C. DNA methylation is a key regulator in transcription and altered effect of methylation behavior has been implicated in many diseases like cancer, atherosclerosis, aging etc. [31,32]. Post-translational modification can occur at any step in the "life cycle" of a protein. For example, many proteins are modified shortly after translation is completed to mediate proper protein folding or stability or to direct the nascent protein to distinct cellular compartments (e.g., nucleus, membrane). Other modifications occur after folding and localization are completed to activate or inactivate catalytic activity or to otherwise influence the biological activity of the protein. Proteins are also covalently linked to tags that target a protein for degradation. Besides single modifications, proteins are often modified through a combination of post-translational cleavage and the addition of functional groups through a step-wise mechanism of protein maturation or activation.

CONCLUSION

Results obtained show a significant genetic alteration ofthe *MED12* gene with a high frequency of mutations noted inparticular codon 44 of exon 2. All these mutations beingpredicted as deleterious testify to their implication in thepathobiology of uterine fibroids. In addition, the notedalterations lead to instability of the MED12 protein and thusa change in itsbiological function in uterine fibroids.

The results obtained thus open up avenues forunderstanding the molecular mechanisms involved in theoccurrence and/or progression of uterine fibroids. They alsoprovide a glimpse of treatment strategies because MED12proves to be an indispensable biomarker in the progression ofuterine fibroids.

EthicalStatements

All procedures on human genetic material and data within this study were performed in accordance with the ethical principles of the local Ethics Committee of the Cheikh Anta Diop University. Informed consent was obtained from all individual participants prior to the inclusion in the study.

REFERENCES

[1] A. Audebert, "External endometriosis: histogenesis, etiologyand natural evolution", Rev Practitioner, vol. 40, 1990, pp.1077-1081.

[2] D. D. Baird, D. B. Dunson, M. C. Hill, D. Cousins, J. M.Schectman, " High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence ", Am J Obstet Gynecol, vol. 188, 2003, pp. 100-107.

[3] S. Okolo, " Incidence, aetiology and epidemiology of uterine fibroids", Best Pract Res Clin Obstet Gynaecol, vol. 22, 2008, pp. 571-588.

[4] N. Mäkinen, M. Mehine, J. Tolvanen, E. Kaasinen, Y. Li, H. J.Lehtonen, M. Gentile, J. Yan, M. Enge, M. Taipale, M.Aavikko, R. Katainen, E. Virolainen, T. Böhling, T. A. Koski,V. Launonen, J. Sjöberg, L. A. Aaltonen, "MED12, the Mediator Complex Subunit 12 Gene, Is Mutated at High Frequency in Uterine Leiomyomas", Science, vol. 334, 2011, pp. 252-254

[5] R. J. Kelleher, P. M. Flanagan, R. D. Kornberg, R. D, A novel Mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell, vol. 61, 1990, pp. 1209-1215.

[6] Y. J. Kim, S. Björklund, Y. Li, M. H. Sayre, and R. D. Kornberg, A multiprotein Mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell, vol. 77, 1994, pp. 599-608.

[7] F. S. Collins, L. D. Brooks, A. A. Chakravarti, DNA polymorphism discovery resource for research on human genetic variation. Genome Res, vol. 8, 1998, pp. 1229-1231.

[8] E. Capriotti, R.B. Altman, Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinform,vol.12, 2011.

[9] M. Petukh, T.G. Kucukkal, E. Alexov, On human disease-causing amino acid variants: statistical study of sequence and structural patterns. Hum. Mutat, vol. 36, 2015, pp. 524-534.

[10] B. Keneme, D. Ciss, S. Ka, A. Dem, M. Sembene, S.M. Gueye. Prediction of the Structure andMutations Instability of the Med12 Exon2 Gene in Uterine Fibroids in Senegalese Women, International Journal of Genetics and Genomics, vol. 7, No. 3, 2019, pp. 80-87.

[11] M.J. Landrum, S. Chitipiralla, G.R. Brown, C. Chen, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Kaur, C. Liu, V. Lyoshin, Z. Maddipatla, R. Maiti, J. Mitchell, N. O'Leary, G.R. Riley, W. Shi, G. Zhou, V. Schneider, D. Maglott, J.B. Holmes, B.L. Kattman, ClinVar: improvements to accessing data. Nucleic Acids Res, vol. 48, 2020, pp. 835-844.

[12] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the Expasy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, 2005, pp. 571-607.

[13] S. Ittisoponpisan, S.A. Islam, T. Khanna, E. Alhuzimi, A. David, M.J.E. Sternberg, Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J. Mol. Biol,vol. 431, 2019, pp. 2197-2212.
[14] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H. Nam, M. Mort, D.N. Cooper, J. Sebat, L.M. Iakoucheva, S.D. Mooney, P. Radivojac, Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat. Commun. vol. 11, 2020, 5918.
[15] P. Radivojac, Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. Proteins, vol. 63, 2006, pp. 398-410.
[16] J. Thusberg, A. Olatubosun, M. Vihinen, Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat., vol. 32, 2001, pp. 358-368.
[17] P. Radivojac, Identification, analysis, and prediction of protein ubiquitination sites. Proteins, vol. 78, 2010, pp. 365-380.
[18] C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Science, vol. 30, 2021, pp. 60-69.
[19] D.E. Pires, D.B. Ascher, T.L. Blundell. mCSM:Predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics, vol. 30, 2014, pp. 335-342.
[20] C.H. Rodrigues, D.E. Pires, D.B. Ascher. DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res, vol. 46, 2018, pp. 350-355.
[21] H.H. Magnus, E.N. Kiehl, B. Petersen, M. Nielsen, O. Winther, H. Nielsen, J. Hallgren, P. Marcatili, NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning, 2022, https://doi.org/10.1093/nar/gkac439
[22] E. M. Je, M. R. Kim, K. O. Min, N. J. Yoo, S. H. Lee,"Mutational analysis of MED12 exon 2 in uterine leiomyomaand other common tumors", Int J Cancer, 2012.
[23] M. A. De Graaff, A. M. Cleton-Jansen, K. Szuhai, J. V. Bovee, «Mediator complex subunit 12 exon 2 mutation analysis indifferent subtypes of smooth muscle tumors confirms geneticheterogeneity" Hum Pathol, vol. 44, No. 8, 2013, pp. 1597-1604.
[24] C. E. Barbieri, S. C. Baca, M. S. Lawrence, "Exomesequencing identifies recurrent SPOP, FOXA1 and MED12mutations in prostate cancer". Nat Genet, vol. 44, No. 6, 2012,pp. 685-689
[25] C. E. Barbieri, A. Sboner, M. A. Rubin, L. A. Garraway, "Mutation in prostate tumours from Caucasian patients", JPathol, vol. 230, 2013, pp. 453-456.
[26] H. M. Bourbon, "Comparative genomics supports a deep evolutionary origin for the large, four module transcriptionalmediator complex". Nucleic Acids Res, vol. 36, 2008, pp.3993-4008
[27] T. M. Knuesel, D. K. Meyer, J. A. Donner, J M. J. Espinosa, J.D. Taatjes, "The Human CDK8 Subcomplex Is a HistoneKinase That Requires Med12 for Activity and Can FunctionIndependently of Mediator ", Mol Cell Biol, vol. 29, No. 3,2009, pp. 650-661
[28] M. Turunen, J. M. Spaeth, S. Keskitalo, "Uterine leiomyoma-linked MED12 mutations disrupt Mediator associated CDKactivity", Cell Rep, vol. 7, No. 3, 2014, pp. 654-660
[29] J.M. Thornton, 1992. Lessons from analyzing protein structures. Curr.Opin. Struct. Biol, vol. 2, pp. 888-894.

[30] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature, vol. 431, 2004, pp. 931-45.
[31] O.N.Jensen, Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol, vol. 8, 2004, pp. 33-41.
[32] J.R. Wagner, S. Busche, B. Ge, T. Kwan,T. Pastinen, M. Blanchette, The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol, vol. 15, 2014, R37.