Available online freely at www.isisn.org

Bioscience Research

Print ISSN: 1811-9506 Online ISSN: 2218-3973

Journal by Innovative Scientific Information & Services Network

RESEARCH ARTICLE BIOSCIENCE RESEARCH, 2022 19(4): 1873-1883. **OPEN ACCESS**

PTEN gene mutational spectrum in a Saudi Arabian cohort with suspected Mendelian diseases: an in-silico variant analysis

Hamsa Jameel Banjer.

Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taif University, **Saudi Arabia**

***Correspondence: h.banjer@tu.edu.sa** **Received 02-09-2022, Revised: 13-10-2022, Accepted: 22-10-2022 e-Published: 24-10-2022**

This study aims to analyze in-silico the most frequent variants of PTEN in a Saudi Arabian cohort with suspected Mendelian disease. PTEN mutation variants data of Saudi cohort extracted from the Saudi Human Genome Project (SHGP) database. Individual samples were sequenced for the PTEN gene. In-silico analysis using the various bioinformatic tools were carried out for pathogenicity and functional predication. the analysis showed that around 5% of mutations had a frequency of more than 40% of homozygous alleles, based on the elevated rate of consanguinity in Saudi Arabia. Interestingly, eight variants were found at >1% of the analyzed cohort and were considered single nucleotide variations (SNV). Compared with other geographical populations, the rs1799734 and rs701848 PTEN variants were found to have increased by around 30% in allele frequencies. Variation found in PTEN allele frequencies could be representing a common Arab-enriched or Arab-specific common variants. Optimistically, this work may contribute to better preventive medicine, timely diagnoses, and an improved understanding of the history of PTEN-related disorders.

Keywords: PTEN; Single nucleotide variation; Variants;, Allele frequencies; Saudi Genome; cancer predisposition

INTRODUCTION

PTEN is a tumor suppressor gene that acts as a dual-specificity phosphatase in several pathways implicated in cellular growth. PTEN dysfunction causes dysregulation of the PI3K/AKT/ mTOR growth-promoting signaling cascade and other pathways, causing hereditary predisposition to cancer and overgrowth disorders. Germline mutations of PTEN can result in a broad-spectrum pf phenotypes, ranging from cancer to autism. Therefore, the PTEN hamartoma tumor syndrome (PHTS) term was cond. PHTS is a highly variable autosomal dominant condition associated with intellectual disability, overgrowth, and phenotypes of tumor predisposition that frequently overlap. PHTS has several historical clinical presentations, including Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), and macrocephaly-autism/developmental delay syndrome (Tan et al. 2012; Macken, Tischkowitz and Lachlan, 2019; Yehia, Keel and Eng, 2020).

PTEN holds nine exons that canonically encode a protein of 403 amino acids (Li *et al.* 1997). Generally, PTEN mutations affect the abundance of the protein, resulting in haplo-insufficiency and reduced or lost phosphatase activity, behaving in a dominant-negative manner, and/or resulting in abnormal localization and function (Lee, Chen and Pandolfi, 2018). Germline mutations spectrum of PHTS is broad, with numerous mutations impacting all of PTEN nine exons. Additionally, various pathogenic promoter mutations were found to influence the transcription and translation of PTEN gene due to the alteration of RNA secondary structure (Zhou et al. 2003a; Heikkinen et al. 2011). Recently, some unsuspected intronic variants of PTEN gene have been reported to result in pathogenic exon skipping, alternative splicing, or the use of cryptic splice sites (Chen *et al.* 2017). These splicing changes were correlated with significant lower levels of PTEN protein and p-AKT elevation in patients with splicing changes compared with those with normal splicing. However, PTEN large deletions were found to occur in approximately 3–10% of PHTS patients over the entire coding sequence (Zhou et al. 2003b; Mester and Eng, 2013; Yehia, Ngeow and Eng, 2019).

Interestingly, genetic disorders occur at a high frequency in several Arab communities. About two-thirds of these genetic diseases in Arab patients are inherited in autosomal recessive mode. High fertility rates, combined with an increasing number of consanguineous marriages (25–60% of all marriages are consanguineous, and first-cousin marriages are common) in Arab populations, tend to increase the rates of genetic and congenital abnormalities (Abedalthagafi, 2019; AlHarthi et al. 2020). Many of the nearly 500 genes explored in Arab individuals have revealed remarkably heterogeneous spectra, and

several novel and rare mutations have caused many different clinical outcomes (Tadmouri, Sastry and Chouchane, 2014). Population-specific studies on genetic variants and diversity have provided information on variants, such as the type of variation, the location of the variation in a chromosome, the relevant genes, and population-specific allele frequencies and genotypes, all of which are critically important in precision medicine and individualized prevention and management. This study aimed to identify PTEN genetic variants present in a Saudi Arabian cohort as well as their in-silico analysis.

MATERIALS AND METHODS

2.1 Data

Data on PTEN gene variant mutations were obtained from the Saudi Human Genome Project (SHGP) database in March 2021. The data were drawn from a Saudi population cohort that was analyzed between 2013 and 2015. According to the SHGP database, from a Saudi cohort of 6,038, individual samples (screenee) were sequenced for the PTEN gene on the date of data extraction. The readings were mapped to the UCSC GRCh37/hg19 assembly (Church et al. 2011). SHGP is an ongoing project that aims to sequence 100,000 Saudi genomes. To date, 56,799 genome samples have been sequenced (https://shgp.kacst.edu.sa/). Authorization on the use of SHGP database information for publication is approved upon citation of the following work (Saintenac, Access and Saintenac, 2015; Abouelhoda, Sobahy, et al. 2016). All sequenced individuals were enrolled under several King Faisal Specialist Hospital and Research Centre institutional review board-approved protocols after providing an assigned informed written consent form relevant to the disease with which they presented (Saintenac, Access and Saintenac, 2015; Abouelhoda, Faquih, et al. 2016; Abouelhoda, Sobahy, et al. 2016). These individuals were phenotyped for suspected Mendelian diseases. Venous blood was collected for DNA extraction as mentioned on the SHGP publications (Saintenac, Access and Saintenac, 2015; Abouelhoda, Faquih, et al. 2016; Abouelhoda, Sobahy, et al. 2016). The study was conducted in accordance with the Helsinki declaration.

Moreover, UCSC genome browser, ClinVar, GenomAD, dbSNP and Kaviar (Glusman et al. 2011) databases were used to identify various data on the variants (e.g., variant positions, allele frequencies, reported conditions, etc.).

2.2 Sequence Variant Nomenclature

The Mutalyzer 2 Position Converter and Name Generator was used to convert chromosomal positions to transcript-oriented positions and to generate valid variant descriptions, respectively, based on the guidelines of the Human Genome Variation Society (HGVS). Mutalyzer 2 checks sequence variant nomenclature according to the HGVS guidelines (den Dunnen et al. 2016; Lefter et al. 2021a).

2.3 In-silico predictions

The variants of the PTEN gene identified in this study were classified using the in-silico variant pathogenicity predictor MutationTaster2 (Steinhaus et al. 2021). For the PTEN g.89690828G>A variant, BayesDel (addAF and noAF), DEOGEN2, EIGEN/EIGEN PC, FATHMM, LIST-S2, LRT, M-CAP, MVP, Mutation assessor, PROVEAN, PrimateAI and SIFT were used through their own webservers(Ionita-Laza et al. no date; Chun and Fay, 2009; Kumar, Henikoff and Ng, 2009; Choi, 2012; Choi et al. 2012; Shihab et al. 2013; Jagadeesh et al. 2016; Raimondi et al. 2017; Sundaram et al. 2018; Malhis et al. 2020; Qi et al. 2021). Each method yields a quantitative prediction score measuring the likelihood of a variant being deleterious or not. Some qualitative predictions were assigned based on the specific cutoff score of deleteriousness endorsed by the creators, as shown in Table 6. MetaLR, MetaSV, MetaRNN, and REVEL Meta Score algorithm predictors were used to determine pathogenicity based on the combined evidence from multiple other in-silico predictors (Liu, Jian and Boerwinkle, 2011; Dong et al. 2015a; Liu et al. 2020; Li et al. 2021). Moreover, protein functional domains were identified through the InterPro database (Blum et al. 2021). DynaMut predictors were used to assess the impact of mutations on protein stability and dynamics (Rodrigues, Pires and Ascher, 2018).

RESULTS

3.1 PTEN variant dataset from dbSHGP

This retrospective study investigated PTEN gene variants. Datasets of PTEN gene variants were retrieved from the SHGP database. The data extraction yielded a cohort of 6,038 individuals (screened). Samples were analyzed for PTEN genes by DNA sequencing, either through WGS or targeted gene panel analysis (Saintenac, Access and Saintenac, 2015; Abouelhoda, Sobahy, *et al.* 2016). Sequencing of PTEN in this cohort showed a total of 292 different variants, or 4.84% of the analyzed samples. The zygosity of the variants revealed the following: 41.3% homozygous and 69.8% heterozygous variant allele frequency (VAF). Single nucleotide variants in more than 1% of the studied population were found at 2.47%, while the rest of the variants were considered less than 1% of the investigated population. The types of mutation varied as follows: 3.59% base substitution, 0.59% insertion, and 0.64% deletion. A summary of these results is provided in Table 1.

Table 1: Summery of PTEN variants.

PTEN gene		Count	Percentage % *
PTEN Variants		**292**	**4.84**
Variant allele frequency(VAF)	Homozygous	49	41.3
	Heterozygous	277	69.8
Single-nucleotide variant(SNV)	>1%	8	2.74
	<1%	284	97.26
Mutation types	Substitution	217	3.59
	Insertion	36	0.60
	Deletion	39	0.65

* Percentages were calculated relevant to the cohort sample numbers.

3.2 Single nucleotide variation (SNV)

PTEN variants of dbSHGP were sorted based on variant frequency. The 30 most frequent variants were selected for further analysis (Table 2). Moreover, dbSNP identifiers (IDs) for most of the variants were identified, except for seven variants that were not identified by the dbSNP or by any other genomic variant database, as verified by Kaviar (Glusman et al. 2011) and other SNV search tools. As table 2 shows, eight PTEN variants were found at >1%, and the rest of the variant frequencies were found below 1%. Notably, the first two variants were present at a high frequency in the investigated population, at 45% and 33%, respectively. Furthermore, the first eight variants were considered SNP because they were found in >1% of the analyzed cohort. Thus, compared with other reported population frequencies described on the reference dbSNP IDs, variants 1 (rs555895) and 2 (rs1903858) were at similar frequencies. Variants 4 (rs71022512) , 5 (rs2943772) and 7 (rs34003473) showed an elevated allele frequencies by 5% to 1% of increment. Variant 6 (rs1120260) showed around 5% less on mutation frequencies. Interestingly, variants 3 (rs1799734) and 8 (rs701848) were shown to be almost 30% lower than the reported allele frequencies reported on their dbSNP IDs. Furthermore, the types of nucleotide variations were mainly single nucleotide exchanges, in addition to six variants of the indel mutation.

Table 2: Variant frequencies of the top 30 PTEN mutations

	Chromosomal variant *	Nucleotide change	Variant Frequency	Variant Frequency (Homo)	Variant Frequency (Hetero)	SNP IDs*
1	g.89720907T>G	T/G	0.451	0.160	0.290	rs555895
2	g.89653686A>G	A/G	0.336	0.120	0.216	rs1903858
3	g.89690955_89690959insTCTTA	- / TCTTA	0.047	0.019	0.028	rs1799734
4	g.89623861delT	T/-	0.047	0.047	0.000	rs71022512
5	g.89623901G>C	G/C	0.047	0.047	0.000	rs2943772
6	g.89727414C>T	C/T	0.016	0.001	0.014	rs1120260
7	g.89720634delT	T/-	0.015	0.000	0.015	rs34003473
8	g.89726745T>C	T/C	0.012	0.004	0.007	rs701848
9	g.89623323G>A	G/A	0.009	0.001	0.008	rs1044322
10	g.89711855_89711856insT	- / T	0.007	0.000	0.007	rs756623620
11	g.89725294delT	T/-	0.007	0.000	0.007	---
12	g.89685398G>A	G/A	0.005	0.000	0.005	rs185262832
13	g.89624377A>G	A/G	0.005	0.000	0.005	---
14	g.89720633delT	T/-	0.005	0.000	0.005	rs376702513
15	g.89624116G>T	G/T	0.004	0.000	0.003	rs761148721
16	g.89653719A>T	A/T	0.003	0.000	0.003	rs185005124
17	g.89624135C>T	C/T	0.003	0.000	0.003	---
18	g.89712152G>A	G/A	0.002	0.001	0.001	rs149607752
19	g.89623716G>A	G/A	0.002	0.000	0.002	rs12573787
20	g.89623581C>T	C/T	0.002	0.000	0.002	rs866865188
21	g.89717799G>A	G/A	0.002	0.000	0.002	rs116160352
22	g.89712028A>G	A/G	0.002	0.000	0.002	---
23	g.89720633C>T	C/T	0.002	0.000	0.002	rs376702513
24	g.89624131A>T	A/T	0.002	0.001	0.000	---
25	g.89624132A>T	A/T	0.002	0.001	0.001	---
26	g.89711834delT	- / T	0.002	0.000	0.002	---
27	g.89624133G>A	G/A	0.001	0.001	0.000	---
28	g.89690828G>A	G/A	0.001	0.000	0.001	rs202004587
29	g.89624340C>T	C/T	0.001	0.000	0.001	rs190707033
30	g.89720580T>A	T/A	0.001	0.000	0.001	rs1207532534

*Some of the dbSNP Identifiers are equivalence to genomic location of the Gh38 assembly.

3.3 Mutation analysis

All 30 variants were analyzed in-silico using the Mutation Taster tool for functional prediction. Benign mutations were predicted for almost all analyzed variants (Table 3). Interestingly, one of the results predicted a deleterious effect on variant No. 27: g.89720580T>A. Thus, Random Forest classification algorithm of this tool results ranged from 0 –25 decision trees for deleterious alterations, and 100–299 decision trees of benign alterations. Notably, regarding the g.89720580T>A variant, 89 decision trees showed deleterious alterations, and only 11 decision trees showed benign alterations. Almost all mutational changes were positioned on the intronic regions and UTRs. More importantly, all the variants analyzed using MutationTaster2 were predicted to have no effect on splicing sites. However, the indel variants were not compatible with MutationTaster2 analysis.

3.4 Intronic variants

Table 4 presents the coding DNA references obtained by the Mutalyzer Position Converter using the genomic location (Lefter et al. 2021b). The detection of variant gene regions by the genome viewer positioned gene alterations on different introns of the PTEN 8 introns, while some other variants were positioned on the UTR at the first and last exons of the 9 PTEN exons, as mentioned earlier (Figure 1). Additionally, one mutation was found in the flanking region of 5'UTR. Moreover, most mutations were not detected at any splicing site. The findings of several recent studies have suggested that deep intronic variants, that is, variants that lie >100 bp away from the nearest exon–intron junctions, might lead to a variety of significant effects on gene expression. Notably, 11 variants were considered deep intronic variants because they were located more than 100 base pairs away from exon–intron junctions (Table 4).

Table 3: Prediction of mutation effect

	Chromosomal variant *	SNP IDs*	Prediction	Tree vote (del | benign)	Gene Region	Splice site change
1	g.89720907T>G	rs555895	Benign	0|100	Intron	No
2	g.89653686A>G	rs1903858	Benign	0|100	Intron	No
3	g.89690955_89690959insTCTTA	rs1799734	---	---	---	---
4	g.89623861delT	rs71022512	---	---	---	---
5	g.89623901G>C	rs2943772	Benign	3|297	5'UTR	No
6	g.89727414C>T	rs1120260	Benign	0|100	3'UTR	No
7	g.89720634delT	rs34003473	-	-	-	-
8	g.89726745T>C	rs701848	Benign	0|100	3'UTR	No
9	g.89623323G>A	rs1044322	Benign	2|298	5'UTR	No
10	g.89711855_89711856insT	rs756623620	---	---	---	---
11	g.89725294delT	---	---	---	---	---
12	g.89685398G>A	rs185262832	Benign	0|100	Intron	No
13	g.89624377A>G	---	Benign	0|100	Intron	No
14	g.89720633delT	rs376702513	Benign	25|275	Intron	No
15	g.89624116G>T	rs761148721	---	---	---	---
16	g.89653719A>T	rs185005124	Benign	0|100	Intron	No
17	g.89624135C>T	---	Benign	25|275	5'UTR	No
18	g.89712152G>A	rs149607752	Benign	0|100	Intron	No
19	g.89623716G>A	rs12573787	Benign	21|279	5'UTR	No
20	g.89623581C>T	rs866865188	Benign	7|293	5'UTR	No
21	g.89717799G>A	rs116160352	Benign	0|100	Intron	No
22	g.89712028A>G	---	Benign	0|100	Intron	No
23	g.89720633C>T	rs376702513	Benign	0|100	Intron	No
24	g.89624131A>T	---	Benign	1|299	5'UTR	No
25	g.89624132A>T	---	Benign	1|299	5'UTR	No
26	g.89711834delT	---	---	---	---	---
27	g.89624133G>A	---	Benign	18|282	5'UTR	No
28	g.89690828G>A	rs202004587	Deleterious	89|11	CDS	No
29	g.89624340C>T	rs190707033	Benign	0|100	Intron	No
30	g.89720580T>A	rs1207532534	Benign	0|100	Intron	No

*Some of the dbSNP Identifiers are equivalence to genomic location of the Gh38 assembly.

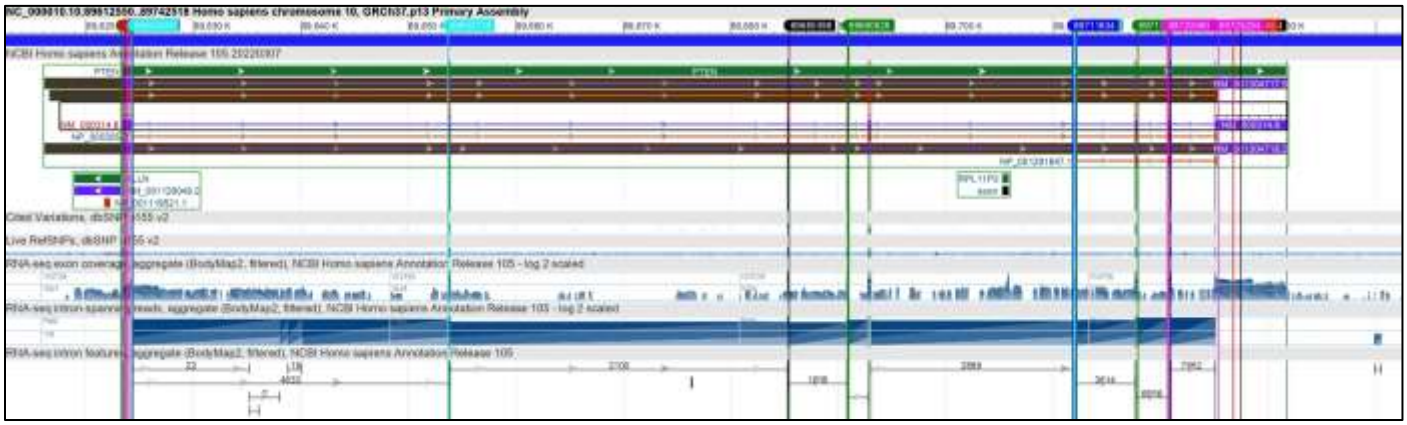

Figure 1: Locations of the top 30 variants on PTEN Gene. Mapping of the variants were done using genomic positions as indicated by the colored vertical lines

Table 4: PTEN intronic variants.

	Chromosomal variant	Coding Variant[1]	Alteration region	No. of Exon / Intron	Position[2]	Deep Intronic Variant
1	g.89720907T>G	c.1026+32T>G	Intron	Intron 8	32	<100bp
2	g.89653686A>G	c.80-96A>G	Intron	Intron 1	-96	<100bp
3	g.89690955_89690959insTCTTA	c.253+109_253+113insTCTTA	Intron	Intron 4	109-253	>100bp
4	g.89623861delT	c.-366delT	5'UTR	Exon 1	-366	>100bp
5	g.89623901G>C	c.-326G>C	5'UTR	Exon 1	-326	>100bp
6	g.89727414C>T	c.*2185C>T	3'UTR	Exon 9	*2185	>100bp
7	g.89720634delT	c.802-17delT	Intron	Intron 7	-17	<100bp
8	g.89726745T>C	c.*1516T>C	3'UTR	Exon 9	*1516	>100bp
9	g.89623323G>A	c.-904G>A	5'UTR	Upstream Intron 1	-904	>100bp
10	g.89711855_89711856insT	c.493-20_493-19insT	intron	Intron 5	- 20 to -19	>100bp
11	g.89725294delT	c.*65delT	3'UTR	Exon 9	*65	<100bp
12	g.89685398G>A	c.209+84G>A	Intron	Intron 3	84	<100bp
13	g.89624377A>G	c.79+72A>G	Intron	Intron 1	72	<100bp
14	g.89720633delT	c.802-18delT	Intron	Intron 7	-18	<100bp
15	g.89624116G>T	c.-111G>T	5'UTR	Exon 1	-111	>100bp
16	g.89653719A>T	c.80-63A>T	Intron	Intron 1	-63	<100bp
17	g.89624135C>T	c.-92C>T	5'UTR	Exon 1	-29	<100bp
18	g.89712152G>A	c.634+136G>A	Intron	Intron 6	136	>100bp
19	g.89623716G>A	c.-511G>A	5'UTR	Exon 1	-511	>100bp
20	g.89623581C>T	c.-646C>T	5'UTR	Exon 1	-646	>100bp
21	g.89717799G>A	c.801+23G>A	Intron	Intron 7	23	<100bp
22	g.89712028A>G	c.634+12A>G	Intron	Intron 6	12	<100bp
23	g.89720633C>T	c.802-18C>T	Intron	Intron7	-18	<100bp
24	g.89624131A>T	c.-96A>T	5'UTR	Exon 1	-96	<100bp
25	g.89624132A>T	c.-95A>T	5'UTR	Exon 1	-95	<100bp
26	g.89711834delT	c.493-41delT	Intron	Intron 5	-41	<100bp
27	g.89624133G>A	c.-94G>A	5'UTR	Exon 1	-94	<100bp
28	g.89690828G>A	c.235G>A	CDS	Exon 4	235	-
29	g.89624340C>T	c.79+35C>T	Intron	Intron 1	35	<100bp
30	g.89720580T>A	c.802-71T>A	Intron	Intron 7	-71	<100bp

[1] Sequence variants including the (plus) (minus) and (asterisk) are described according to the HGVS-nomenclature. [2] Distance from the closest exon-intron boundary.

3.5 Indel variants

Most prediction tools using genomic position coordinates have been designed for the analysis of single nucleotide changes. Thus, the indel variants of the top 30 PTEN variants were checked for their effects and pathogenicity using ClinVar and previously published reports (Table 5). Similarly, all indel variants indicated benign or likely benign effects under conditions of PHTS, hereditary cancer-predisposing syndrome, and CS

3.5 Indel variants

Most prediction tools using genomic position coordinates have been designed for the analysis of single nucleotide changes. Thus, the indel variants of the top 30 PTEN variants were checked for their effects and

pathogenicity using ClinVar and previously published reports (Table 5). Similarly, all indel variants indicated benign or likely benign effects under conditions of PHTS, hereditary cancer-predisposing syndrome, and CS.

3.6 chr10:g.89690828G>A variant

The variant g.89690828G>A (c.235G>A) is a missense variant located on exon 4, with a variation that causes amino acid substitution, that is, p.A79T. However, the dbSNP database classifies this variant as of uncertain significance, while ClinVar classifies it as likely benign (dbSNP: rs202004587, ClinVar: 41682). Additionally, several conditions linked to this variant, which were reported from 2012–2022, indicated its clinical impact as benign, likely benign, or of uncertain significance. The reported conditions varied between PHTS, CS, macrocephaly-autism syndrome, and hereditary cancer-predisposing syndrome. Moreover, in almost all the indicated conditions, this variant was identified as a germline mutation characterized by autosomal dominant inheritance. Notably, this mutation was observed at a 1.4% frequency of homozygous mutation in the SHGP cohort (Table 2). Moreover, the pathogenicity of this variant was predicted using various algorithms, as shown in Table 6.

Table 5: Indels variants of PTEN gene IDs and pathogenicity.

Variant No.	Chromosomal variant	SNP ID	ClinVar interpretation	Condition
3	g.89690955_89690959 ins TCTTA	rs1799734	Not Reported in ClinVar	---
4	g.89623861delT	rs71022512	Benign	Not specified
7	g.89720634delT	rs34003473	Benign/Likely benign	PHTS Hereditary cancer-redisposing syndrome
10	g.89711855_89711856insT	rs756623620	Benign/Likely benign	PHTS, CS Hereditary cancer-redisposing syndrome
11	g.89725294delT	rs878853930	Benign	PHTS
15	g.89624116G>T	rs761148721	Uncertain significance	PHTS, CS
26	g.89711834delT	---	Not Reported in ClinVar	---

*Some of the dbSNP Identifiers are equivalence to genomic location of the Gh38 assembly.

Table 6: In-silico predictions of the PTEN g.89690828G>A variant.

Algorithmic method		Prediction	Score	Rank score	Range[1]
Meta Scores	MetaLR	Damaging	0.7996	0.9322	0 to 1
	MetaSVM	Damaging	0.3912	0.8896	-2.0058 to +3.0399
	MetaRNN	Tolerated	0.3464	0.5159	0 to 1
	REVEL	Benign	0.485	0.7765	0 to 1
Individual Predictions	BayesDel addAF	Damaging	0.08652	0.6279	-1.29334 to 0.75731
	BayesDel noAF	Damaging	0.1462	0.7994	-2.0058 to +3.0399.
	DEOGEN2	Tolerated	0.4127	0.7668	---
	EIGEN	Benign	-0.0983	0.3746	---
	EIGEN PC	Benign	0.1291	0.4604	---
	FATHMM	Damaging	-4.93	0.9839	-16.13 to +10.64
	FATHMM-MKL	Damaging	0.671	0.3331	0 to 1
	FATHMM-XF	Damaging	0.6021	0.5941	---
	LIST-S2	Damaging	0.8753	0.5863	---
	LRT	Deleterious	0.000024	0.5587	0 to 1
	M-CAP	Damaging	0.09697	0.7669	0 to 1
	MVP	Pathogenic	0.981	0.9808	---
	Mutation assessor	Neutral	-0.39	0.03026	-5.135 to +6.49
	PROVEAN	Neutral	0.09	0.05917	-14 to +14
	PrimateAI	Damaging	0.8159	0.8441	---
	SIFT	Tolerated	0.579	0.06011	0 to 1
	SIFT4G	Tolerated	0.594	0.08106	0 to 1

[1] Ranges with empty boxes are assigned as cut off-value.

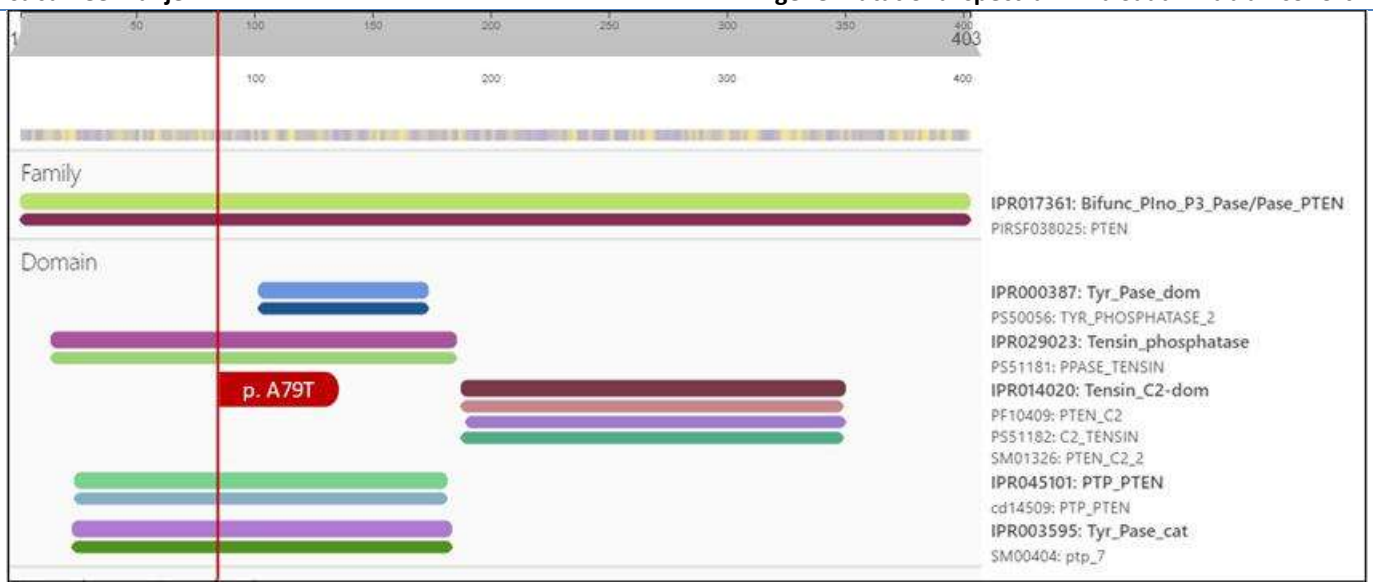

Figure 2: PTEN Protein domains. The red line indicates the p.A79T variant site at PTEN amino acid sequence.

The in-silico analysis of the variant using the meta-scores algorithm revealed a damaging effect at a score of 0.7996 by the MetaLR algorithm and a score of 0.3912 by the MetaSV algorithm (Dong et al. 2015b). In contrast, the MetaRNN and REVEL algorithms predicted tolerant and benign scores of 0.3464 and 0.485, respectively. Individual prediction scores are shown in Table 6. Furthermore, a protein domain check by InterPro revealed that a p.A79T mutation was located on the catalytic tensin-type phosphatase domain (amino acids 14–185) (Figure 2). The DynaMut prediction suggested that the effects of the A79T substitution on the conformational dynamics and stability of the PTEN protein destabilize it (ΔΔG: −0.3431 kcal/mol; a negative value of ΔΔG indicates the mutation destabilizes the protein). Additionally, the GO annotation to the PTEN protein indicated a dephosphorylation biological process (GO:0016311) and phosphatase activity molecular function (GO:0016791).

DISCUSSION

This study explored the mutational spectrum of PTEN tumor suppressor gene variants in a Saudi Arabian cohort drawn from the SHGP. Germline and somatic mutations in PTEN can cause a spectrum of clinical syndromes, including CS, BRRS, autism spectrum disorders with macrocephaly, and PHTS. Thus, the significance of this gene warrants a context-dependent investigation, which could lead to better variomics research and improved diagnosis, personalized prevention, and treatment practices.

Summarized counts of the PTEN variants in the SHGP dataset demonstrated that around 5% of mutations had a frequency of more than 40% of homozygous alleles, likely due to the elevated rate of consanguinity in Saudi Arabia. Interestingly, eight variants were found at >1% of the analyzed cohort and were considered SNV. Compared with other geographical populations described in reference dbSNP IDs, variants 3 (rs1799734) and 8 (rs701848) were found to be increased by around 30% in allele frequencies. This finding may represent Arab-specific or Arab-enriched common variants. Moreover, this finding highlights the importance of anthropologically informed sampling in researching human genetic diversity, especially for personalized and disease-oriented preventive medicine.

The in-silico functional prediction of the top 30 PTEN variants indicated that almost all mutations were benign. Additionally, it was found that these mutations were located on the intronic and UTR regions, away from any splicing sites. These findings could be indicating that these mutations are not disease-causing, as introns and UTR used to be considered noncoding DNA sequences. However, it has been proposed that noncoding regions may affect gene function and expression (Chen et al. 2017; Qian et al. 2021). In addition, previous studies on mis-splicing mutations were based on exome data and were mainly limited to exons and canonical splice sites (Jung, Lee and Choi, 2021). Notably, in the present study, 11 PTEN variants were considered deep intronic mutations. The presence of this type of intronic mutations called for attention to the mutation *per se* and its carriers, as many mutations are mostly context dependent. A previous study reported that various PTEN intronic mutations near splice junctions can result in exon skipping, premature termination, or a shift in isoform usage (Chen et al. 2017). However, to the best of our knowledge, comprehensive studies of intronic mutations have not been conducted because of the lack of large-scale matched whole-genome sequencing and RNA-seq data.

The overall functional prediction of p.A79T revealed contradictory predictions regarding its pathogenicity.

Multiple in silico algorithms predicted a pathogenic variant, while others suggested a benign effect on protein function. However, p.A79T was found to be located on the functional domain responsible for its phosphatase activity. Thus, it is important to maintain observations of PTEN variant carriers because the mutation was present in 1.4% of the studied population. Moreover, this percentage is higher than that of previously reported allele frequencies, in which this variant was found at a frequency of 0.008% in healthy individuals and a frequency of 0.02% in individuals and families with CS, BRRS, PHTS, Lynch syndrome, or breast and ovarian cancer (Figer et al. 2002; Ngeow et al. 2011; Pilarski et al. 2011; Tan et al. 2011, 2012; Nizialek et al. 2015; Yurgelun et al. 2015; Dominguez-Valentin et al. 2018; Momozawa et al. 2018)

CONCLUSION

In conclusion, variomics provides the foundation for personalized medicine by linking genetic variations to disease expression, outcome, and treatment, yet its utility depends on appropriate assays to evaluate the effects of mutations on protein. In this study, an in-silico analysis was conducted on the spectrum of the PTEN variant. The studied cohort, who was drawn from the SGHP, illustrated the different PTEN variants mutations and their frequencies in Saudi Arabia population. Also, it revealed a wide range of intronic mutation variants with elevated allele frequencies and differing zygosity. In fact, these information and data call for greater epidemiologic studies considering the evolving role of the mutations in prognosis and personalized management.

Based on the outcomes of this study, variomics studies importance lies relevance of discovering the hidden diversity of variants in different geographical regions to aid in the genetic diagnostic yield. Improved knowledge of mutations and variants could lead to the development of preventive measures to reduce the likelihood of developing genetic-related disorders.

CONFLICT OF INTEREST

The authors declared that present study was performed in absence of any conflict of interest.

ACKNOWLEDGEMENT

The author wishes to thank all colleagues for their valuable and constructive suggestions during the planning and development of this research work.

AUTHOR CONTRIBUTIONS

H.B. conceptualized the study, collected the data, carried out all the analysis and wrote and revised the manuscript.

REFERENCES

Abedalthagafi, M. (2019) 'Precsion medicine of monogenic disorders: Lessons learned from the Saudi human Genome', Frontiers in Bioscience - Landmark, 24(5), pp. 870–889. Available at: https://doi.org/10.2741/4757.

Abouelhoda, M., Faquih, T., El-Kalioby, M. and Alkuraya, F.S. (2016) 'Revisiting the morbid genome of Mendelian disorders', Genome Biology, 17(1), pp. 1–7. Available at: https://doi.org/10.1186/s13059-016-1102-1.

Abouelhoda, M., Sobahy, T., El-Kalioby, M., Patel, N., Shamseldin, H., Monies, D., Al-Tassan, N., Ramzan, K., Imtiaz, F., Shaheen, R. and Alkuraya, F.S. (2016) 'Clinical genomics can facilitate countrywide estimation of autosomal recessive disease burden', Genetics in Medicine, 18(12), pp. 1244–1249. Available at: https://doi.org/10.1038/gim.2016.37.

AlHarthi, F.S., Qari, A., Edress, A. and Abedalthagafi, M. (2020) Familial/inherited cancer syndrome: a focus on the highly consanguineous Arab population. Available at: https://doi.org/10.1038/s41525-019-0110-y.

Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A. and Finn, R.D. (2021) 'The InterPro protein families and domains database: 20 years on', Nucleic Acids Research, 49(D1), pp. D344–D354. Available at: https://doi.org/10.1093/NAR/GKAA977.

Chen, H.J., Romigh, T., Sesock, K. and Eng, C. (2017) 'Characterization of cryptic splicing in germline PTEN intronic variants in Cowden syndrome', Human Mutation, 38(10), pp. 1372–1377. Available at: https://doi.org/10.1002/humu.23288.

Choi, Y. (2012) 'A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein', 2012 ACM Conference on Bioinformatics, Computational Biology and Biomedicine, BCB 2012, pp. 414–417. Available at: https://doi.org/10.1145/2382936.2382989.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. (2012) 'Predicting the Functional Effect of Amino Acid Substitutions and Indels', PLOS ONE, 7(10), p.

e46688. Available at: https://doi.org/10.1371/JOURNAL.PONE.0046688.

Chun, S. and Fay, J.C. (2009) 'Identification of deleterious mutations within three human genomes', Genome Research, 19(9), pp. 1553–1561. Available at: https://doi.org/10.1101/GR.092619.109.

Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., Albracht, D., Kremitzki, M., Rock, S., Kotkiewicz, H., Kremitzki, C., Wollam, A., Trani, L., Fulton, L., Fulton, R., Matthews, L., Whitehead, S., Chow, W., Torrance, J., Dunn, M., Harden, G., Threadgold, G., Wood, J., Collins, J., Heath, P., Griffiths, G., Pelan, S., Grafham, D., Eichler, E.E., Weinstock, G., Mardis, E.R., Wilson, R.K., Howe, K., Flicek, P. and Hubbard, T. (2011) 'Modernizing reference genome assemblies', PLoS biology, 9(7). Available at: https://doi.org/10.1371/JOURNAL.PBIO.1001091.

Dominguez-Valentin, M., Evans, D.G.R., Nakken, S., Tubeuf, H., Vodak, D., Ekstrøm, P.O., Nissen, A.M., Morak, M., Holinski-Feder, E., Martins, A., Møller, P. and Hovig, E. (2018) 'Genetic variants of prospectively demonstrated phenocopies in BRCA1/2 kindreds', Hereditary Cancer in Clinical Practice, 16(1). Available at: https://doi.org/10.1186/S13053-018-0086-0.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015a) 'Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies', Human Molecular Genetics, 24(8), pp. 2125–2137. Available at: https://doi.org/10.1093/HMG/DDU733.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015b) 'Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies', Human Molecular Genetics, 24(8), pp. 2125–2137. Available at: https://doi.org/10.1093/HMG/DDU733.

den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., Mcgowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E. and Taschner, P.E.M. (2016) 'HGVS Recommendations for the Description of Sequence Variants: 2016 Update', Human Mutation, 37(6), pp. 564–569. Available at: https://doi.org/10.1002/humu.22981.

Figer, A., Kaplan, A., Frydman, M., Lev, D., Paswell, J., Papa, M.Z., Goldman, B. and Friedman, E. (2002) 'Germline mutations in the PTEN gene in Israeli patients with Bannayan–Riley–Ruvalcaba syndrome and women with familial breast cancer', Clinical Genetics, 62(4), pp. 298–302. Available at: https://doi.org/10.1034/J.1399-0004.2002.620407.X.

Glusman, G., Caballero, J., Mauldin, D.E., Hood, L. and Roach, J.C. (2011) 'Kaviar: an accessible system for testing SNV novelty', Bioinformatics, 27(22), pp. 3216–3217. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR540.

Heikkinen, T., Greco, D., Pelttari, L.M., Tommiska, J., Vahteristo, P., Heikkilä, P., Blomqvist, C., Aittomäki, K. and Nevanlinna, H. (2011) 'Variants on the promoter region of PTEN affect breast cancer progression and patient survival', Breast Cancer Research, 13(6), pp. 1–11. Available at: https://doi.org/10.1186/BCR3076/FIGURES/3.

Ionita-Laza, I., Mccallum, K., Xu, B. and Buxbaum, J. (no date) 'A SPECTRAL APPROACH INTEGRATING FUNCTIONAL GENOMIC ANNOTATIONS FOR CODING AND NONCODING VARIANTS'.

Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A. and Bejerano, G. (2016) 'M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity', Nature genetics, 48(12), pp. 1581–1586. Available at: https://doi.org/10.1038/NG.3703.

Jung, H., Lee, K.S. and Choi, J.K. (2021) 'Comprehensive characterisation of intronic mis-splicing mutations in human cancers', Oncogene 2021 40:7, 40(7), pp. 1347–1361. Available at: https://doi.org/10.1038/s41388-020-01614-3.

Kumar, P., Henikoff, S. and Ng, P.C. (2009) 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm', Nature Protocols 2009 4:7, 4(7), pp. 1073–1081. Available at: https://doi.org/10.1038/nprot.2009.86.

Lee, Y.R., Chen, M. and Pandolfi, P.P. (2018) 'The functions and regulation of the PTEN tumour suppressor: new modes and prospects', Nature reviews. Molecular cell biology, 19(9), pp. 547–562. Available at: https://doi.org/10.1038/S41580-018-0015-0.

Lefter, M., Vis, J.K., Vermaat, M., den Dunnen, J.T., Taschner, P.E.M. and Laros, J.F.J. (2021a) 'Next Generation HGVS Nomenclature Checker', Bioinformatics (Oxford, England), 37(18), pp. 2811–2817. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTAB051.

Lefter, M., Vis, J.K., Vermaat, M., den Dunnen, J.T., Taschner, P.E.M. and Laros, J.F.J. (2021b) 'Next Generation HGVS Nomenclature Checker', Bioinformatics (Oxford, England), 37(18), pp. 2811–2817. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTAB051.

Li, C., Zhi, D., Wang, K. and Liu, X. (2021) 'MetaRNN: Differentiating Rare Pathogenic and Rare Benign Missense SNVs and InDels Using Deep Learning', bioRxiv, p. 2021.04.09.438706. Available at: https://doi.org/10.1101/2021.04.09.438706.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang,

S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S.H., Giovanella, B.C., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M.H. and Parsons, R. (1997) 'PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer', Science (New York, N.Y.), 275(5308), pp. 1943–1947. Available at: https://doi.org/10.1126/SCIENCE.275.5308.1943.

Liu, X., Jian, X. and Boerwinkle, E. (2011) 'dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions', Human Mutation, 32(8), pp. 894–899. Available at: https://doi.org/10.1002/HUMU.21517.

Liu, X., Li, C., Mou, C., Dong, Y. and Tu, Y. (2020) 'dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs', Genome Medicine, 12(1), pp. 1–8. Available at: https://doi.org/10.1186/S13073-020-00803-9/FIGURES/4.

Macken, W.L., Tischkowitz, M. and Lachlan, K.L. (2019) 'PTEN Hamartoma tumor syndrome in childhood: A review of the clinical literature', American journal of medical genetics. Part C, Seminars in medical genetics, 181(4), pp. 591–610. Available at: https://doi.org/10.1002/AJMG.C.31743.

Malhis, N., Jacobson, M., Jones, S.J.M. and Gsponer, J. (2020) 'LIST-S2: taxonomy based sorting of deleterious missense mutations across species', Nucleic Acids Research, 48(W1), pp. W154–W161. Available at: https://doi.org/10.1093/NAR/GKAA288.

Mester, J. and Eng, C. (2013) 'When overgrowth bumps into cancer: the PTEN-opathies', American journal of medical genetics. Part C, Seminars in medical genetics, 163C(2), pp. 114–121. Available at: https://doi.org/10.1002/AJMG.C.31364.

Momozawa, Y., Iwasaki, Y., Parsons, M.T., Kamatani, Y., Takahashi, A., Tamura, C., Katagiri, T., Yoshida, T., Nakamura, S., Sugano, K., Miki, Y., Hirata, M., Matsuda, K., Spurdle, A.B. and Kubo, M. (2018) 'Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls', Nature Communications 2018 9:1, 9(1), pp. 1–7. Available at: https://doi.org/10.1038/s41467-018-06581-8.

Ngeow, J., Mester, J., Rybicki, L.A., Ni, Y., Milas, M. and Eng, C. (2011) 'Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-like syndrome characterized by germline PTEN, SDH, or KLLN alterations', The Journal of clinical endocrinology and metabolism, 96(12). Available at: https://doi.org/10.1210/JC.2011-1616.

Nizialek, E.A., Mester, J.L., Dhiman, V.K., Smiraglia, D.J. and Eng, C. (2015) 'KLLN epigenotype-phenotype associations in Cowden syndrome', European journal of human genetics : EJHG, 23(11), pp. 1538–1543. Available at: https://doi.org/10.1038/EJHG.2015.8.

Pilarski, R., Stephens, J.A., Noss, R., Fisher, J.L. and Prior, T.W. (2011) 'Predicting PTEN mutations: An evaluation of cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome clinical features', Journal of Medical Genetics, 48(8), pp. 505–512. Available at: https://doi.org/10.1136/jmg.2011.088807.

Qi, H., Zhang, H., Zhao, Y., Chen, C., Long, J.J., Chung, W.K., Guan, Y. and Shen, Y. (2021) 'MVP predicts the pathogenicity of missense variants by deep learning', Nature Communications 2021 12:1, 12(1), pp. 1–9. Available at: https://doi.org/10.1038/s41467-020-20847-0.

Qian, X., Wang, J., Wang, M., Igelman, A.D., Jones, K.D., Li, Y., Wang, K., Goetz, K.E., Birch, D.G., Yang, P., Pennesi, M.E. and Chen, R. (2021) 'Identification of Deep-Intronic Splice Mutations in a Large Cohort of Patients With Inherited Retinal Diseases', Frontiers in Genetics, 12.

Raimondi, D., Tanyalcin, I., FertCrossed, J.S.D., Gazzo, A., Orlando, G., Lenaerts, T., Rooman, M. and Vranken, W. (2017) 'DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins', Nucleic Acids Research, 45(Web Server issue), p. W201. Available at: https://doi.org/10.1093/NAR/GKX390.

Rodrigues, C.H.M., Pires, D.E.V. and Ascher, D.B. (2018) 'DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability', Nucleic acids research, 46(W1), pp. W350–W355. Available at: https://doi.org/10.1093/NAR/GKY300.

Saintenac, C., Access, O. and Saintenac, C. (2015) 'Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases', Genome Biology, 16(1), p. 134. Available at: https://doi.org/10.1186/s13059-015-0693-2.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M. and Gaunt, T.R. (2013) 'Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models', Human mutation, 34(1), pp. 57–65. Available at: https://doi.org/10.1002/HUMU.22225.

Steinhaus, R., Proft, S., Schuelke, M., Cooper, D.N., Schwarz, J.M. and Seelow, D. (2021) 'MutationTaster2021', Nucleic Acids Research, 49(W1), pp. W446–W451. Available at: https://doi.org/10.1093/NAR/GKAB266.

Sundaram, L., Gao, H., Padigepati, S.R., McRae, J.F., Li, Y., Kosmicki, J.A., Fritzilas, N., Hakenberg, J., Dutta, A., Shon, J., Xu, J., Batzloglou, S., Li, X. and Farh, K.K.H. (2018) 'Predicting the clinical impact of human mutation with deep neural networks', Nature genetics, 50(8), p. 1161. Available at: https://doi.org/10.1038/S41588-018-0167-Z.

Tadmouri, G.O., Sastry, K.S. and Chouchane, L. (2014) 'Arab gene geography: From population diversities to

personalized medical genomics', Global Cardiology Science & Practice, 2014(4), p. 394. Available at: https://doi.org/10.5339/GCSP.2014.54.

Tan, M.H., Mester, J., Peterson, C., Yang, Y., Chen, J.L., Rybicki, L.A., Milas, K., Pederson, H., Remzi, B., Orloff, M.S. and Eng, C. (2011) 'A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands', American journal of human genetics, 88(1), pp. 42–56. Available at: https://doi.org/10.1016/J.AJHG.2010.11.013.

Tan, M.H., Mester, J.L., Ngeow, J., Rybicki, L.A., Orloff, M.S. and Eng, C. (2012) 'Lifetime cancer risks in individuals with germline PTEN mutations', Clinical cancer research : an official journal of the American Association for Cancer Research, 18(2), pp. 400–407. Available at: https://doi.org/10.1158/1078-0432.CCR-11-2283.

Yehia, L., Keel, E. and Eng, C. (2020) 'The Clinical Spectrum of PTEN Mutations', Annual Review of Medicine, 71(1), pp. 103–116. Available at: https://doi.org/10.1146/annurev-med-052218-125823.

Yehia, L., Ngeow, J. and Eng, C. (2019) 'PTEN-opathies: from biological insights to evidence-based precision medicine', The Journal of Clinical Investigation, 129(2), pp. 452–464. Available at: https://doi.org/10.1172/JCI121277.

Yurgelun, M.B., Allen, B., Kaldate, R.R., Bowles, K.R., Judkins, T., Kaushik, P., Roa, B.B., Wenstrup, R.J., Hartman, A.R. and Syngal, S. (2015) 'Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients with Suspected Lynch Syndrome', Gastroenterology, 149(3), p. 604. Available at: https://doi.org/10.1053/J.GASTRO.2015.05.006.

Zhou, X.P., Waite, K.A., Pilarski, R., Hampel, H., Fernandez, M.J., Bos, C., Dasouki, M., Feldman, G.L., Greenberg, L.A., Ivanovich, J., Matloff, E., Patterson, A., Pierpont, M.E., Russo, D., Nassif, N.T. and Eng, C. (2003a) 'Germline PTEN Promoter Mutations and Deletions in Cowden/Bannayan-Riley-Ruvalcaba Syndrome Result in Aberrant PTEN Protein and Dysregulation of the Phosphoinositol-3-Kinase/Akt Pathway', The American Journal of Human Genetics, 73(2), pp. 404–411. Available at: https://doi.org/10.1086/377109.

Zhou, X.P., Waite, K.A., Pilarski, R., Hampel, H., Fernandez, M.J., Bos, C., Dasouki, M., Feldman, G.L., Greenberg, L.A., Ivanovich, J., Matloff, E., Patterson, A., Pierpont, M.E., Russo, D., Nassif, N.T. and Eng, C. (2003b) 'Germline PTEN promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant PTEN protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway', American journal of human genetics, 73(2), pp. 404–411. Available at: https://doi.org/10.1086/377109.