Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

PORTLAND PRESS

Research Article

Computational analysis of Cyclin D1 gene SNPs and association with breast cancer

Ayesha Aftab[1,*], Ranjha Khan[2,*], Wasim Shah[2], Muhammad Azhar[2], Ahsanullah Unar[2], Hafiz Muhammad Jafar Hussain[3] and Ahmed Waqas[4]

[1]Department of Biological Sciences, International Islamic university, Islamabad 44000, Pakistan; [2]The First Affiliated Hospital of USTC, Hefei National Laboratory for Physical Sciences at Microscale, The CAS Key Laboratory of Innate Immunity and Chronic Diseases, School of Life Sciences, CAS Centre for Excellence in Molecular Cell Science, Collaborative Innovation Center of Genetics and Development, University of Science and Technology of China, Hefei 230027, China; [3]Shanghai Jiao Tong University, School of Medicine, Shanghai, China; [4]Department of Zoology, Division of Science and Technology, University of Education Lahore, Multan Campus, Multan, Pakistan

Correspondence: Hafiz Muhammad Jafar Hussain (jafarustc@mail.ustc.edu.cn) or Ahmed Waqas (ahmed.waqas@ue.edu.pk)

CCND1 encodes for Cyclin D1 protein and single-nucleotide polymorphisms (SNPs) can modulate its activity. In the present study, the impact of *CCND1* SNPs on structure and/or function of Cyclin D1 protein using *in silico* tools was investigated. Our analysis revealed only one splice site SNP (c.1988+5G<A) can effect *CCND1* function. Subsequently, 78 out of 169 missense variants were predicted as pathogenic by Polyphen2, SIFT, PROVEAN, SNPs&GO, and PANTHER, and 4/78 missense SNPs were further evaluated because these four SNPs were found to be reside in highly conserved region of Cyclin D1. However, they did not show any major impact on tertiary structure and domain of Cyclin D1 but overall R15S and A190S has displayed a significant diseased phenotype and an altered molecular mechanism predicted by MutPred, FATHMM, SNPeffect, SNAP2, and PredictSNP. Consistently, A190S, R179L, and R15S may also cause a decrease in stability of Cyclin D1 anticipated by I-Mutant, HOPE and SNP effect. Furthermore, the Kaplan–Meier plotter has explained that high expression of *CCND1* is associated with less survival rate of breast cancer patients. Altogether our study suggests that c.1988+5G<A, R15S, R179L, and A190S SNPs could directly or indirectly destabilize Cyclin D1.

Introduction

Breast cancer is a heterogeneous type of cell carcinoma with high rate of morbidity and mortality in women [1]. Since 2008, every year approximately 2 million cases of breast cancer are being diagnosed and approximately 50% cases belonged to developing countries with high rate of mortality [2]. Similarly, due to dwindling resources, lack of high-throughput and innovative technologies to deal the breast cancer management and diagnosis is the major reason of continuously increasing cases of breast cancer in developing countries. On the other hand, it has been reported that breast cancer cases are also going to increase in young women. The evidence has demonstrated that women with age <45 years are facing the leading cause of breast cancer [3]. Thus, despite of emergence of new medical approaches and intensive research, still breast cancer is a major health problem and top priority is given to breast cancer in medical research.

CCND1 encodes Cyclin D1 protein that is an important regulator of G1 phase of the cell cycle. Generally, Cyclin D1 function in association with its Cyclin-dependent kinase (CDK) partner such as CDK4 and CDK6, thus, mediating phosphorylation and inactivation of retinoblastoma protein [4]. Dysregulation of Cyclin D1 is frequently linked with various type of cancer in human with diverse histological origin, and thus, it is considered a potential biomarker for diagnosing of different cancers [5,6]. Previous studies have demonstrated that Cyclin D1 overexpression is the main cause of cancer due to the splice modulation by a polymorphism, A870G, in the donor region of the exon 4/intron boundary [7]. Recently, A870G, polymorphism have been reported in the oesophageal adenocarcinoma [8]. Subsequently, its dysregulation is

*These authors contributed equally to this work.

Received: 29 June 2020
Revised: 05 January 2021
Accepted: 08 January 2021

Accepted Manuscript online: 13 January 2021
Version of Record published: 29 January 2021

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

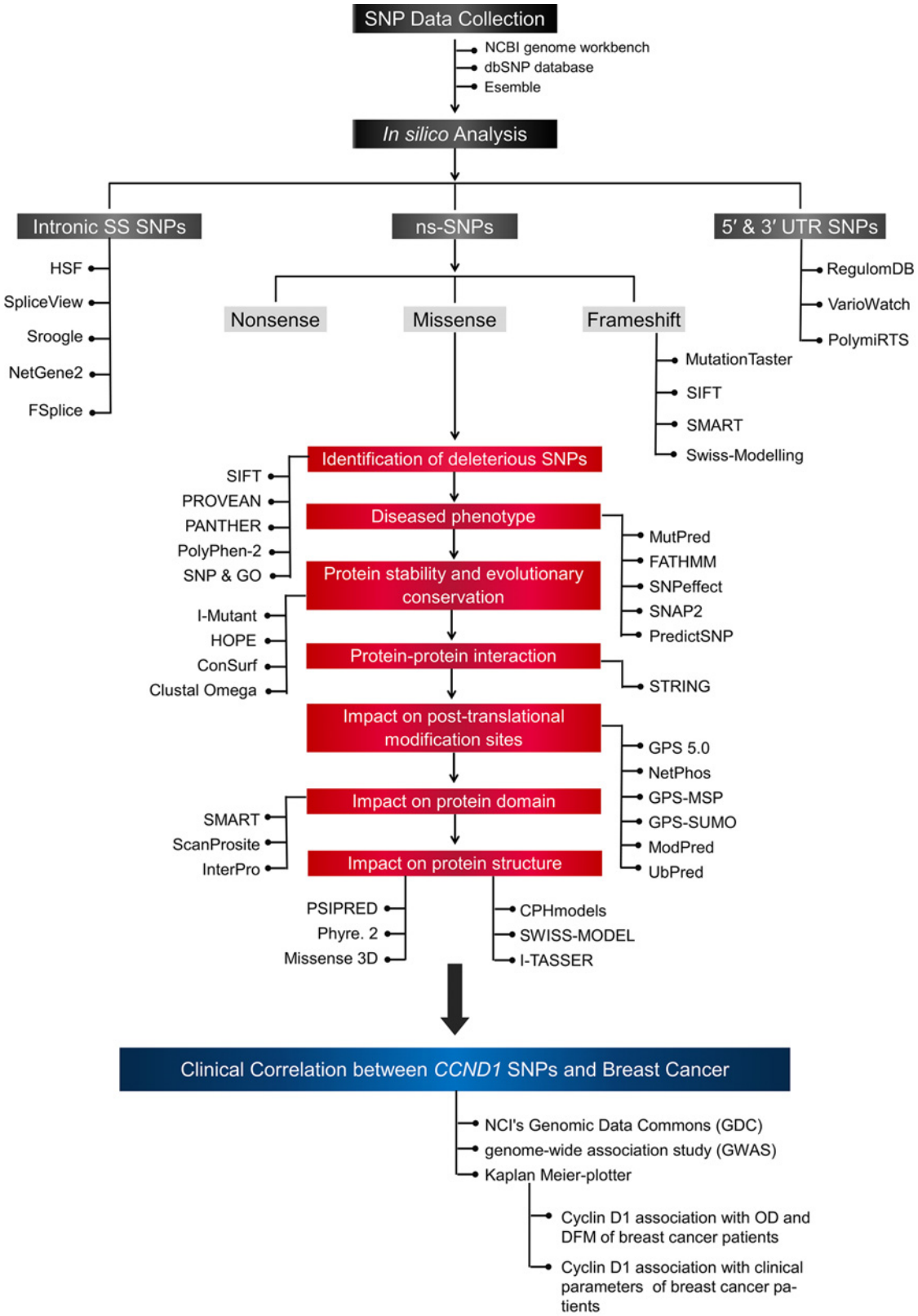


Figure 1. The flow chart of *in silico* analysis steps taken to predict the impact of *CCND1* SNPs on its protein structure and function, and clinical correlation

Abbreviations; DFM, disease-free survival; OS, overall survival; SS, splice site.

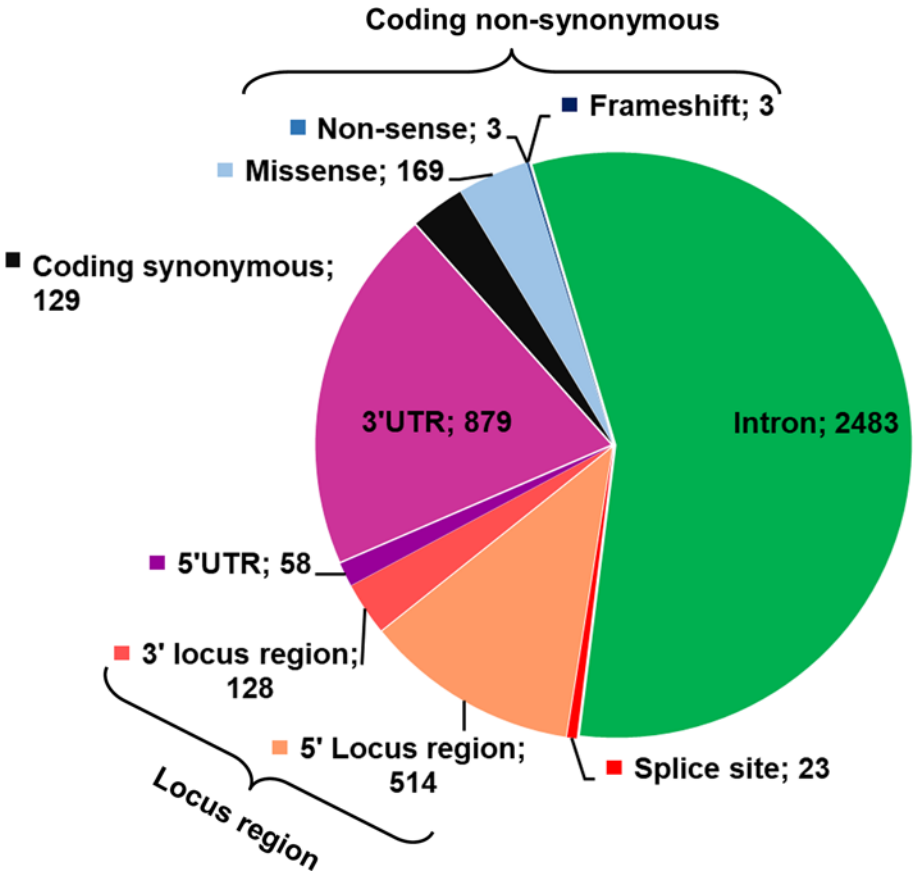


Figure 2. The pie chart displaying the total number of different SNPs of *CCND1*

also reported in breast cancer and transgenic mice of *CCND1* gene also displayed altered mammary cell proliferation and adenocarcinomas [9]. However, the underlying mechanism of Cyclin D1 role in breast cancer is still unknown.

SNPs occur once in every 1000 nucleotides and are positioned in the DNA between genes which are acting as biological marker to locate the genes that are associated with disease. But these SNPs, when occurs within the gene or in regulatory region, they may cause the onset of complex diseases like diabetes and cancer [10]. The Cyclin D1 is found to be involve in complex network signalling with other proteins and forms a *CCND1*–CDK4 complex (DC) with CDK4. The variations in *CCND1* might cause a change in its transcript or translational yield. Therefore, in the present study SNPs of *CCND1* were selected on basis of minor allele frequency (MAF) ranging from 0.0001 to 0.05 and computationally analysed in order to predict their impact on Cyclin D1 function and to evaluate their role in breast cancer.

Methodology

Dataset collection

The *CCND1* SNPs were collected from National Centre for Biotechnology Information (NCBI) genome workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/) and Single Nucleotide Polymorphism Database (dbSNP) of NCBI (https://www.ncbi.nlm.nih.gov/snp). The SNP data of splice site, non-synonymous SNP, 3′ UTR and 5′ UTR SNPs were selected for *in silico* analysis. The details flow chart of steps followed in the present study is given in Figure 1. The nucleotide sequence of *CCND1* and amino acid sequence were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) and ensemble genome browser (https://asia.ensembl.org/index.html), respectively. The UniProt identifier of Cyclin D1 is P24385 and RSC PDB ID is 2w96.A.

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

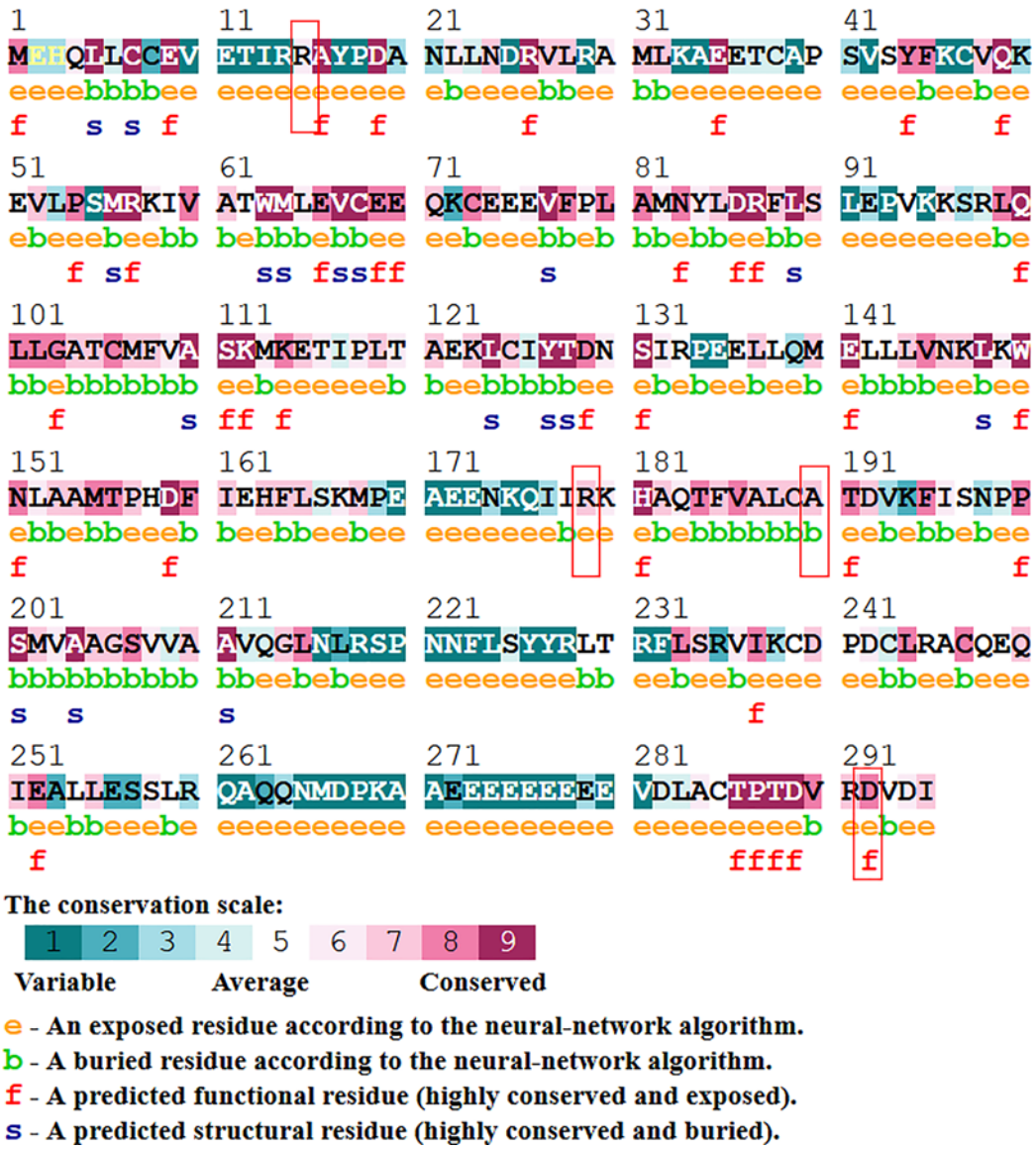


Figure 3. Evolutionary conservation and functional residue prediction by ConSurf

In silico analysis of splice site SNP

Splice site SNPs were selected by considering 10 nucleotides at 5′ and 3′ end of intron. The impact of SNPs on splicing was evaluated by recruiting the five online tools for intronic and splice region mutation. These include the HSF (Human Splicing Finder, http://www.umd.be/HSF3/), SpliceView (http://bioinfo.itb.cnr.it/~webgene/wwwspliceview.html), Sroogle (http://sroogle.tau.ac.il/), Netgene2 (http://www.cbs.dtu.dk/services/NetGene2/), and FSplice v.01 (http://www.softberry.com/berry.phtml?topic=fsplice&group=programs&subgroup=gfind).

In silico analysis of missense variants

The deleterious effect of missense SNPs was analysed using five bioinformatics tools: PolyPhen-2 (Polymorphism Phenotyping v2, http://genetics.bwh.harvard.edu/pph2/), SIFT (Sorting Intolerant from Tolerant, https://sift.bii.a-star.edu.sg/), PROVEAN (Protein Variation Effect Analyzer, http://provean.jcvi.org/index.php), SNP&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html) and PANTHER (http://www.pantherdb.org/tools/). The pathogenic SNPs were selected and additionally filtered and picked on the basis of MAF range 0.0001–0.5, SNP conservation, impact of SNP on protein domain, stability, structure, protein–protein interaction, disease phenotype and post-translational effect. Following tools were accessed for further *in silico* analysis.

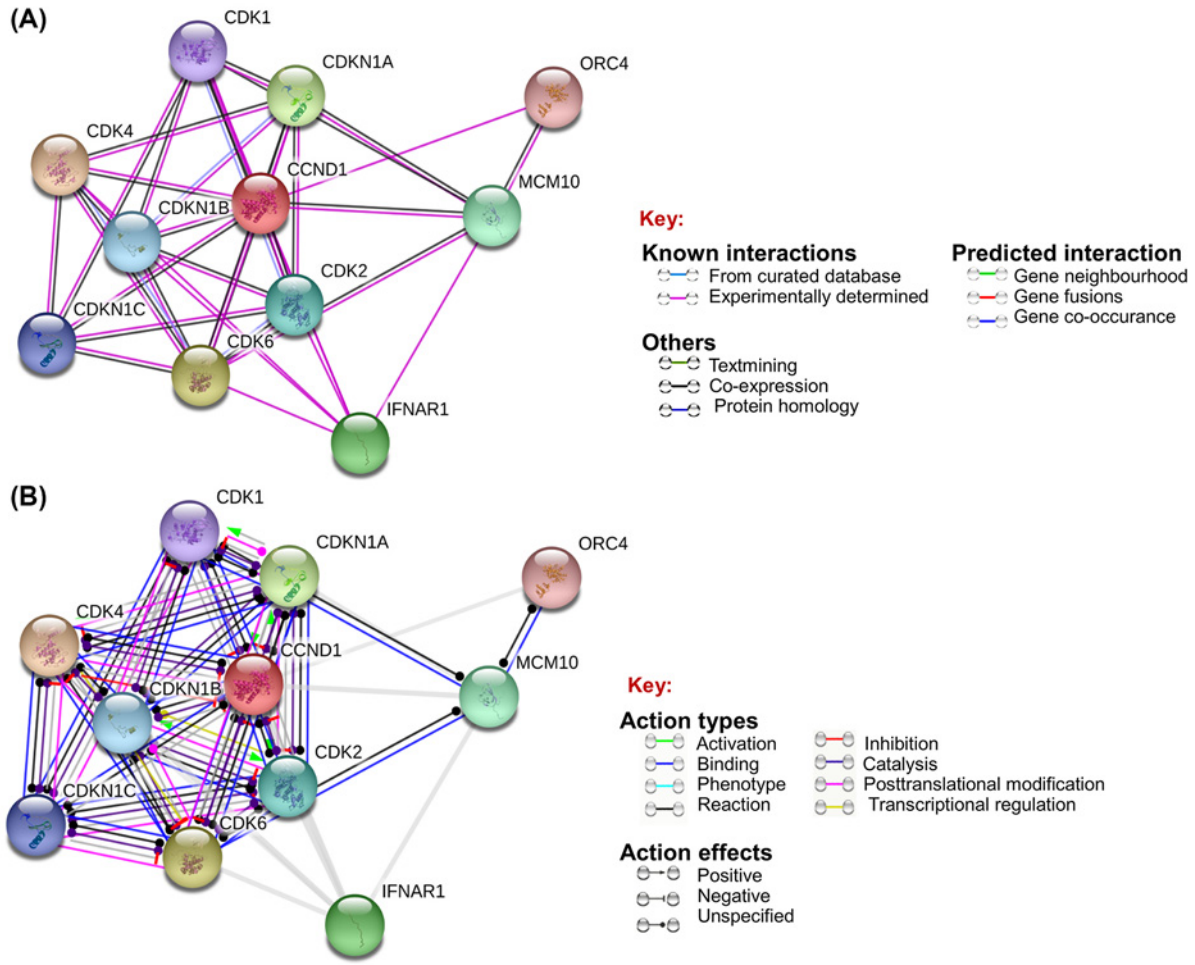


Figure 4. Prediction of protein–protein interaction using STRING v.11.0

(**A**) The interaction evidence and (**B**) molecular action of *CCND1* protein (Cyclin D1) with other proteins.

MutPred (http://mutpred.mutdb.org/) predicts the pathogenicity of amino acid change and the molecular mechanism. It uses Random Forest to predict the g-score and P values to predict the deleterious mutation. The SNP with g score greater the 0.5 are showing the probability of being a deleterious mutation or disease associated. The SNPs are further categorise according to relevant hypotheses. (A) Actionable hypotheses SNP = $g > 0.5$ and $P<0.05$, (B) Confident hypotheses SNP = $g > 0.75$ and $P<0.05$, and (C) Very confident hypotheses SNP = $g > 0.75$ and $P < 0.01$.

FATHMM (Functional Analysis through Hidden Markov Models) v2.3 (http://fathmm.biocompute.org.uk/) is used to predict the functional impact of coding and non-coding variants. CScape option was selected to predict the oncogenic status of four deleterious mutations. The input was given in the form of list having chromosome number, position of variants, and mutant.

SNPeffect database v.4.0 (http://snpeffect.switchlab.org/) was also used to predict the molecular phenotypic impacts SNPs. This database focuses on the effect of mutation on aggregation propensity using TANGO tool, amyloid propensity using WALTZ tool and chaperone binding by using LIMBO tool. It also calculates the effect of mutation on structural stability of protein using FoldX. The input was given as UniProt ID P24385 and corresponding mutation.

SNAP2 (https://rostlab.org/services/snap2web/) is a neural network that distinguishes the effect and neutral SNP by considering the evolutionary conservation, secondary structure, and solvent accessibility effect caused by SNP. The output score ranges from -100 to +100 predicting the strong neutral prediction to strong diseased effect, respectively.

PredictSNP v.2 (https://loschmidt.chemi.muni.cz/predictsnp2/) is a powerful tool that identifies the functional impact of SNP by utilizing six databases CADD, DANN, FATHMM, FitCons, FunSeq2, and GWAVA to develop

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

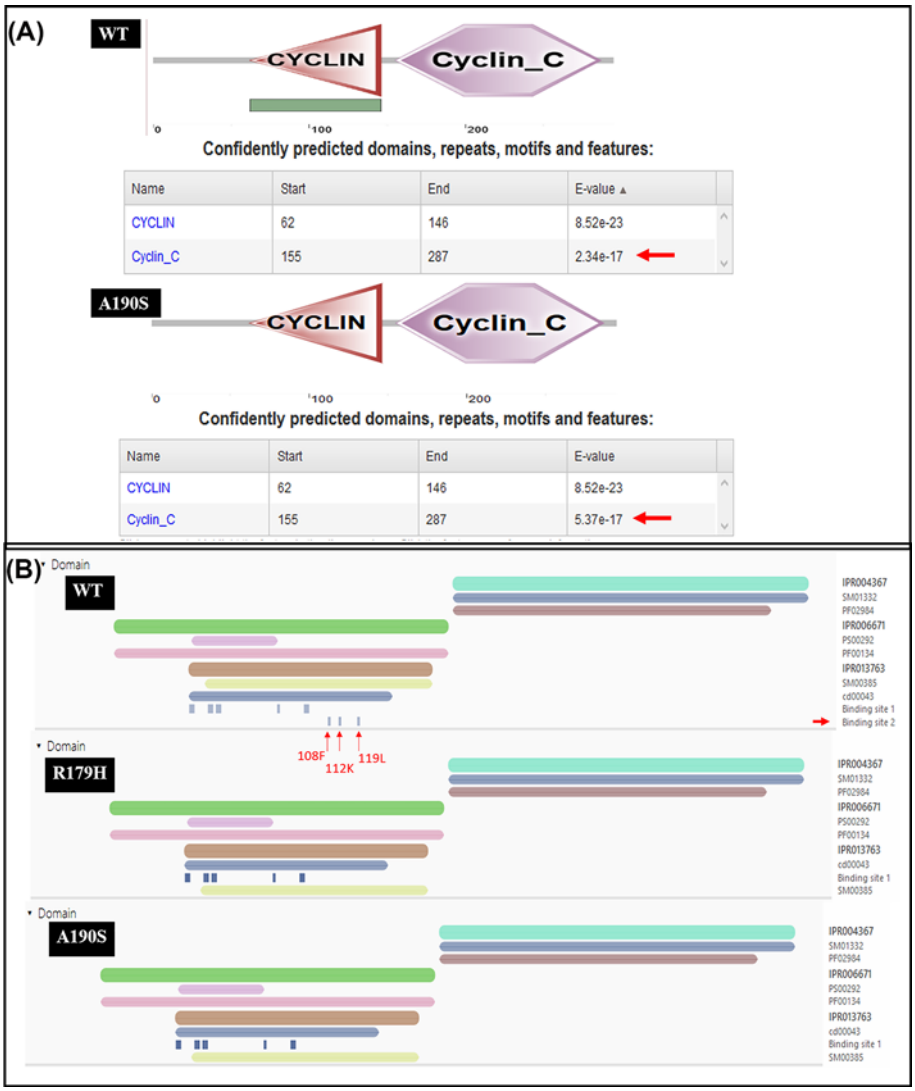


Figure 5. The prediction of SNPs effect on domains of Cyclin D1 using SMART

(**A**) SMART domain analysis showing a change in *e*-value of Cyclin C domain due to A190S SNP. The red arrows indicating the change in *e*-value. (**B**) InterPro analysis showing the missing binding sites due to R179H and A190S SNP. The red arrows indicating the binding sites which become missing from Cyclin D1 protein due to A179S and A90S SNPs.

category-optimal decision thresholds. The output displays the results of five best performing tools in form of neutral, deleterious, and unknown.

HOPE (Have (y)Our Protein Explained, https://www3.cmbi.umcn.nl/hope/) is a next-generation web application for automatic mutant analysis. HOPE combines the information from UniProt, Reprof, and PDB to analyse the effect of mutation on protein structure.

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) is a tool used for prediction of protein stability upon single site mutation. The data set of the tool is resultant from ProTherm which is the most comprehensive database of protein mutation. It predicts the reliability index (RI) ranging from 0 to 10, where 10 is the highest reliability, the DDG in kcal/mol which is the free energy change value [11]. The I-Mutant query was set at 25°C and pH7.

STRING (https://string-db.org/) is a large database of known and predicted protein–protein signalling interactions. The output is given in form of nodes and edges that represent proteins and interaction, respectively. The output scores are indicators of confidence, i.e. how likely STRING judges an interaction to be true, given the available evidence.

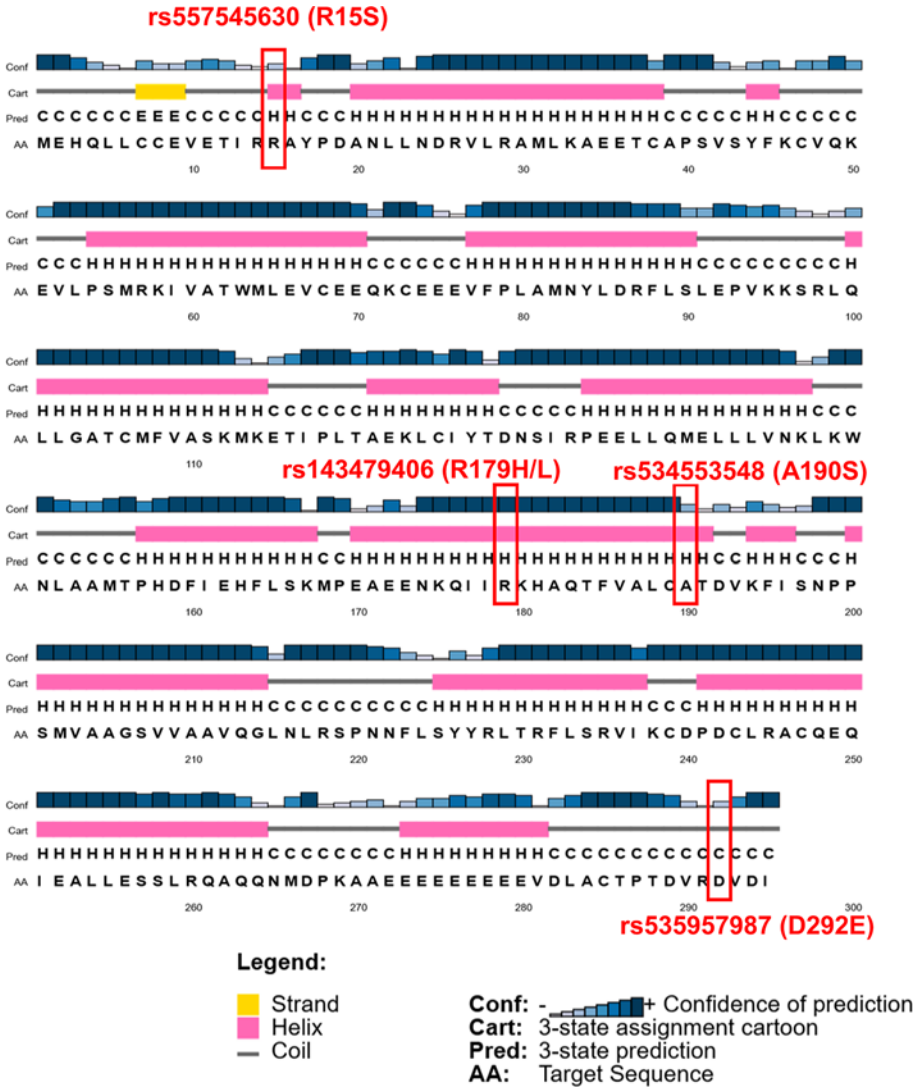


Figure 6. The secondary structure of Cyclin D1 predicted by the PSIPRED

The red rectangles representing the position of SNP.

Instead, they are in order to predict the possible interaction of Cyclin D1 with other proteins; the input name was given as *CCND1* and *Homo sapiens* was selected as organism.

ConSurf Sever (https://consurf.tau.ac.il/) estimates and visualizes evolutionary conservation in macromolecules. The WT amino acid sequence was given as input, HMMER was selected as homology search algorithm, and calculations were based on Bayesian method. According to this method, the conservation score ranges from 1 to 9, the 1–4 is assigned as variable residue, 5–6 as average, and 7–9 as highly conserve residues.

Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) is MSA program for alignment between three or more sequences. The amino acid sequence of human Cyclin D1 (UniProt: P24385) and 19 other species was downloaded, and alignment was performed using online Clustal Omega server.

Prediction of post-translational modifications sites and effect of mutation on it

Post-translational modifications are covalent modifications which modify the protein structure to play an essential role in cellular signalling pathways and networks. For this purpose, we access the easy-to-use CUCKOO Workgroup (http://www.biocuckoo.org/) consisted of several web tools. GPS 5.0 (Group based prediction system), GPS-SUMO, and GPS-MSP (Methyl-group Specific Predictor), BDM-PUB (http://bdmpub.biocuckoo.org/index.php) for prediction of phosphorylation, sumoylation methylation, and ubiquitilation, respectively. In future, we have predicted the

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

Figure 7. PSIPRED sequence plot showing the change in secondary stricture of Cyclin D1 due to R15S SNP

Red arrows indicating the position of SNP. Asterisk (*) representing the location of change of secondary structure in mutant Cyclin D1.

post-translational modification sites in WT and effect of SNP by accessing the NetPhos (http://www.cbs.dtu.dk/services/NetPhos/), ModPred (http://www.modpred.org/) for potential phosphorylation, methylation, sumoylation, and ubiquitilation. Ubpred (http://www.ubpred.org/) was also used for prediction of potential ubiquitination sites in Cyclin D1. To access these tools, the amino acid sequence of Cyclin D1 WT and mutant was submitted in FASTA format.

SNPs impact on Cyclin D1 domains

The impact of SNPs on Cyclin D1 domain was predicted SMART (http://smart.embl-heidelberg.de/), ScanProsite (https://prosite.expasy.org/scanprosite/), and InterPro (https://www.ebi.ac.uk/interpro/). InterPro is an integrated database that predicts and displays the results from different databases like SMART, Prosite, Conserved Domains Database (CCD), and Pfam protein domain database. The input of SNPs in all these tools was given in the form of FASTA format, while WT/ Cyclin D1 domain analysis was done by giving its UniProt ID. The output of WT and mutant protein was compare in order to predict the change caused by SNP. The description of each binding site was taken from InterPro database (https://www.ebi.ac.uk/interpro/).

Secondary structure prediction and structural homology modelling

The effect of SNP on secondary structure of the Cyclin D1 was analysed using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/). Further, Missense3D (http://www.sbg.bio.ic.ac.uk/~missense3d/) was used that matches or compare the WT structure of protein with its mutant and provide detail output of change. Phyre2 (Protein Homology/analogY Recognition Engine V 2.0, http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) predict the secondary structure of the WT protein and mutant protein by providing the FASTA format of the amino acid sequence in input. The homology modelling of the Cyclin D1 and its mutant protein was performed using the SWISS-MODEL (https://swissmodel.expasy.org/), CPHmodels 3.2 server (http://www.cbs.dtu.dk/services/CPHmodels/), and confirmed by I-TASSER (Iterative Threading ASSEmbly Refinement) models (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The structures were visualized and analysed by using Swiss Pdb-Viewer v.4.1.

Frameshift SNPs impact on Cyclin D1 structure and function

CCND1 frameshift mutations were also filtered on the basis of MAF and *in silico* investigation was done to study their pathogenic effect. MutationTaster (http://www.mutationtaster.org/) was accessed to predict the disease-causing

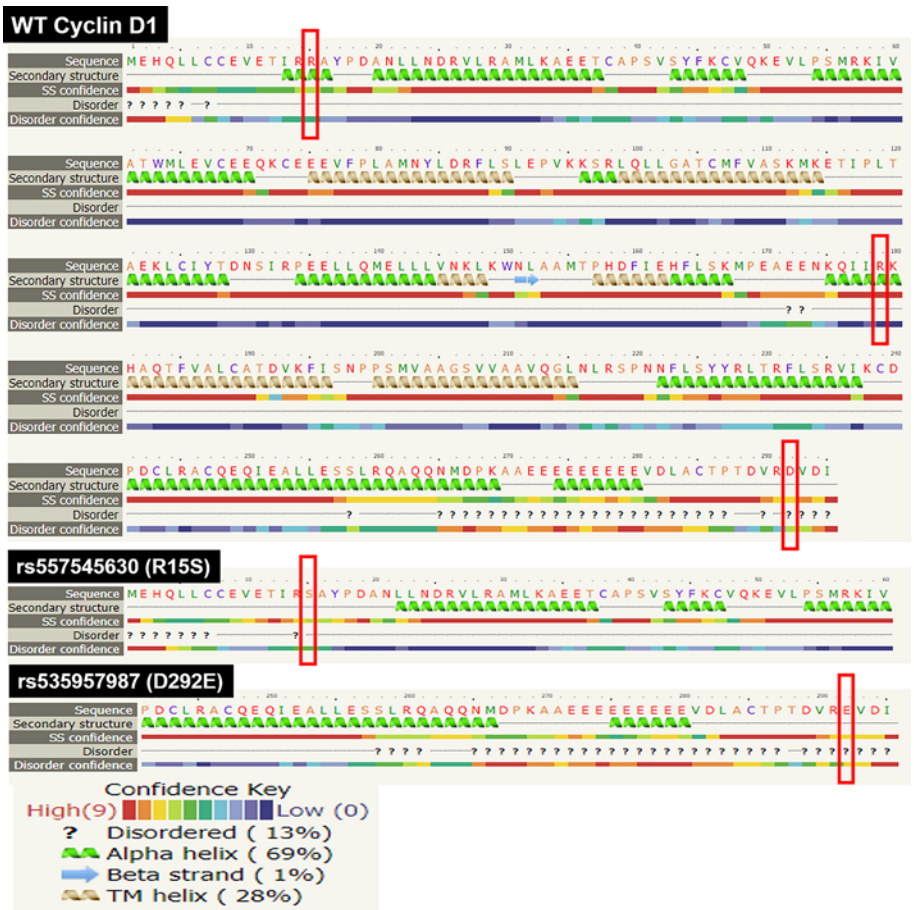


Figure 8. Pyrex secondary structure prediction for WT Cyclin D1 and the two SNPs (R15S and D292E) that have shown effect on secondary structure

The red boxes representing the position of SNPs.

potential effect of an SNP. The impact of SNPs on protein structure and protein domain was predicted by SWISS-MODEL and SMART, respectively. Furthermore, the SIFT *in silico* analysis was performed to verify the pathogenic effect and also the probability of occurrence of non-sense mediated decay (NMD).

Effect of 3′ and 5′ UTR SNPs on Cyclin D1

The 3′ and 5′ UTR SNPs were also selected on the basis of MAF ranges. The effect of SNPs on regulatory units were explored by using multiple *in silico* tools, i.e. RegulomDB (https://regulomedb.org/regulome-search/), Variowatch (http://grch38.genepipe.ncgm.sinica.edu.tw/variowatch/main.do), and PolymiRTS (http://compbio.uthsc.edu/miRSNP/miRSNP_detail_all.php). RegulomDB is a database that predicts regulatory elements in intergenic regions [12]. Variowatch is an automatic data mining tool retrieving genomic information about an SNPs and provides a risk level on functional impact from very low to very high of genomic variants [13]. The PolymiRTS is a database for identified mirSNPs allowing also the evaluation of SNP from dbSNP that classify DNA polymorphism in target sites of miRNA and miRNAs and explain their links physiological behavioural, molecular and disease phenotypes [14].

Clinical association of *CCND1* SNPs

The SNPs which damage the structure and function of a protein may lead to onset of a disease. The *CCND1* SNPs that has shown a clinical association were sorted out using NCBI. The data mining at The NCI's Genomic Data Commons (GDC, https://portal.gdc.cancer.gov/) was performed. GCD is a cancer data repository that provides cancer genomic studies to support precision medicines. We have also performed a comprehensive search to collect the SNPs of *CCND1* that are associated with breast cancer by genome-wide association study (GWAS, https://www.ebi.ac.uk/gwas/). Furthermore, the clinical significance of *CCND1* was related with the survival of breast

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

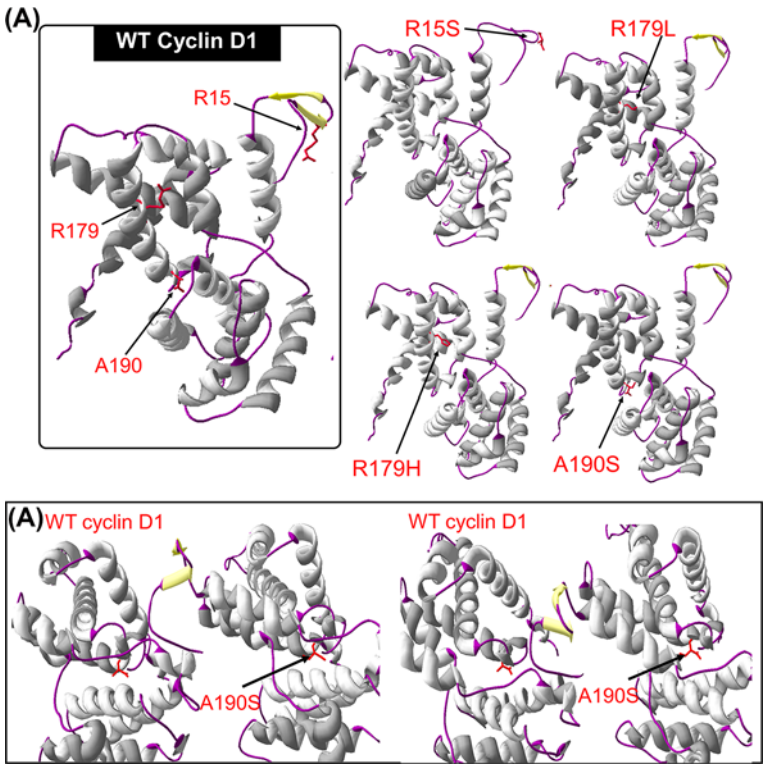


Figure 9. The Swiss-modelling prediction of Cyclin D1 tertiary structure

(**A**) The comparison of WT Cyclin D1 structure with mutant Cyclin D1 due to SNPs. (**B**) The comparison on WT and mutant Cyclin D1 structure due to A190S from different angels, representing the contraction in cavity.

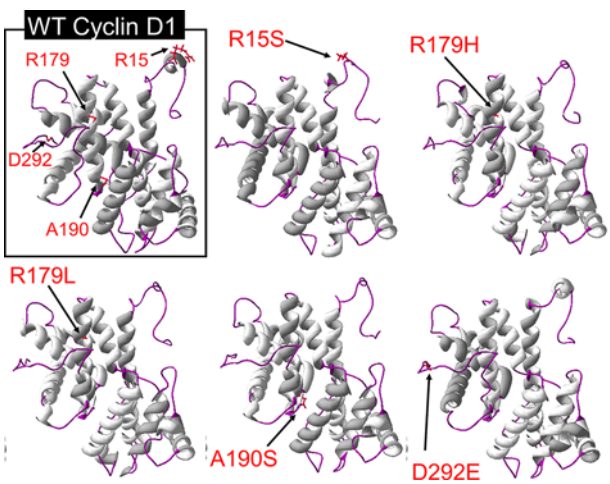


Figure 10. The tertiary structures of WT and mutant Cyclin D1 predicted by I-Tasser

cancer patients using the Kaplan–Meier plotter (https://kmplot.com/analysis/). Kaplan–Meier plotter has assessed of 54,000 genes to evaluate their expression with survival in 21 cancer types and has the data of over 10,000 samples of patients with cancer, out of which 6234 are patients with breast cancer. The source of this system include Gene Expression Omnibus (GEO), European Genome-Phenome Archive (EGA), and The Cancer Genome Atlas (TCGA). In this research study, the *CCND1* expression was also associated with overall survival (OS), disease-free survival (DFS), and clinical parameters of patients with breast cancer.

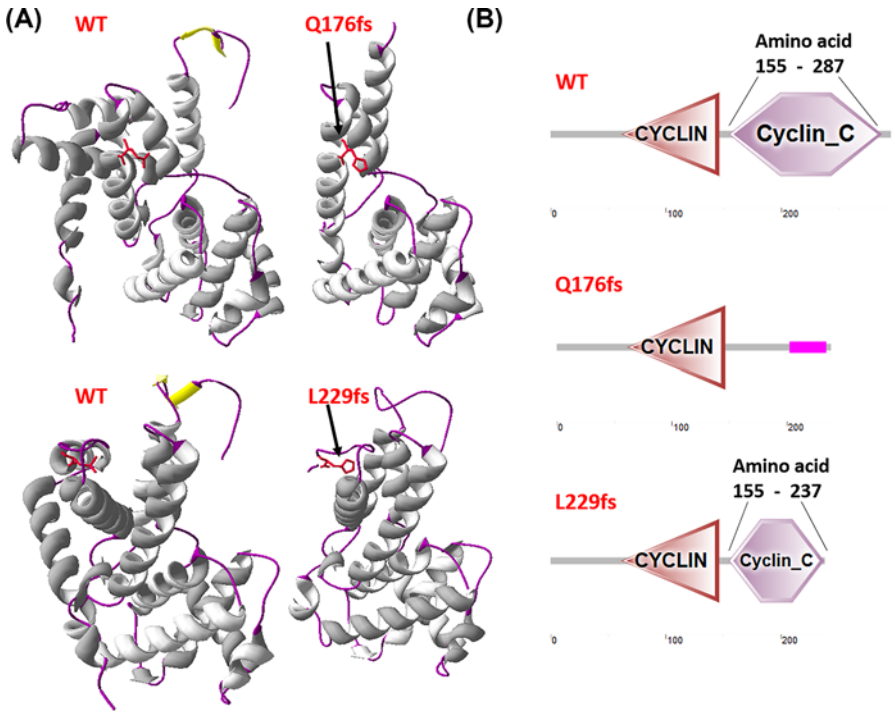


Figure 11. Frameshift SNPs effect on Cyclin D1

(**A**) Prediction of tertiary structure using Swiss-modelling. (**B**) Prediction of domain changes using SMART.

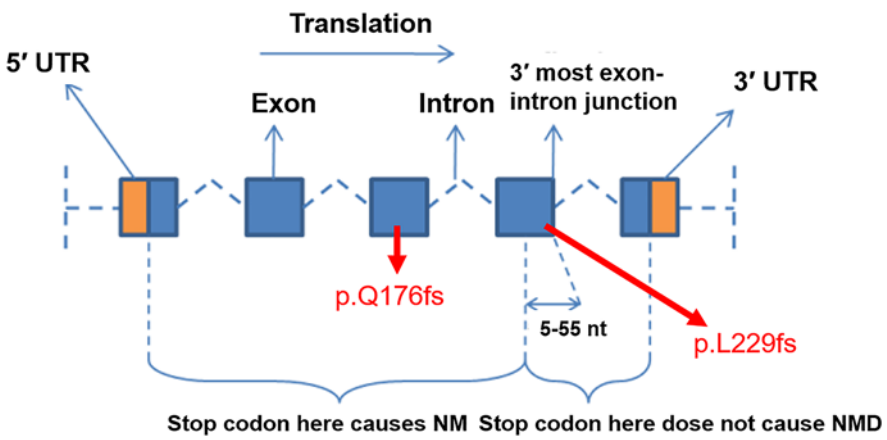


Figure 12. The possible location of occurrence of NMD on Cyclin D1, predicted by SIFT tool analysis

The frameshift SNPs is also positioned on Cyclin D1.

Results

The *CCND1* SNP data were collected from the NCBI genome workbench. Total 3747 *CCND1* SNPs were collected which were consisted of coding non-synonymous SNPs (non-sense, miss-sense and frameshift), coding synonymous SNPs, SNPs in 5′ and 3′ UTR, 3′ and 5′ locus region SNPs and non-coding SNPs (intron and splice site region). The number of different functional classes of SNPs is given in Figure 2. In this research study, the *in silico* analysis of coding non-synonymous SNPs, splice site SNPs and 5′ and 3′ UTR SNP was carried out to predict their pathogenic impact on protein structure and function (Figure 1).

In silico analysis of splice site SNP

The mutation in splicing region causes the improper exon and intron recognition by splicing machinery which results in an aberrant transcript. The variants residing within 10 nucleotide position in intronic splice site region account

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

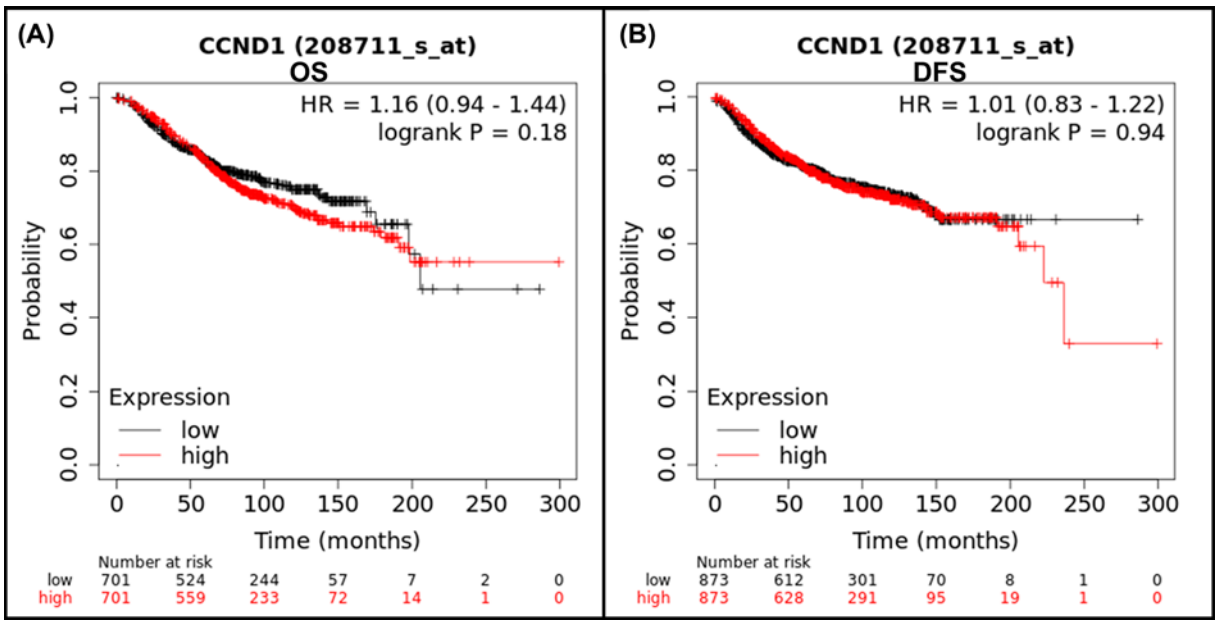


Figure 13. Kaplan–Meier plotter displaying the expression of Cyclin D1 and survival time of patients with breast cancer

(**A**) Overall survival (OS) of patients with breast cancer (*n*=1402). (**B**) Disease-free survival (DFS) of patients with breast cancer (*n*=1746). On the basis of selected parameters, the analysis was run on '*n*' number of patients with available clinical data; HR, hazard ratio.

Table 1 *CCND1* intronic splice site SNP effecting splicing predicted by *in silico* analysis

RS ID	SNP	HSF*matrix score (0–100)				SpliceView score (0–100)		Sroogle			NetGene2 confidence score (0-1)		FSplice (weight index)	
		WT	Mu	Variation %	Interpretation	WT	Mu	Element	WT	Mu	WT	Mu	WT	Mu
rs752676953	g.5412G>A, c.198+5G>A	86.34	74.18	-14.08	Probably effecting splicing	84 DS	78 DS	5′ SS	-3.80 (Delta-G) 6.79 (Max entropy)	-3.80 (Delta-G) 2.02(Max entropy)	0.70 DS	Not generated	W = 7.36	Not generated

*HSF, Human Splicing Finder

more in creating defective splice site [15]. In this research, 23 SNPs out of total 2506 intronic SNPs were collected which occurs within 10 nucleotide position at 5′ and 3′ region of intron. The *in silico* analysis of 23 splice site SNPs was performed by HSF, Sroogle, SliceView, NetGene2 and FSplice comparing the *CCND1* WT and mutant scores. The *in silico* analysis of these tools has predicted that the SNP can affect the donor or acceptor site of splice region (Supplementary Table S1). Out of these 23 SNPS, only one SNP (rs752676953, c.1988+5G<A) was found to be probably damaging by four *in silico* tools, the HSF, Sroogle, NetGene2 and FSplice, given in Table 1. The Netgene2 and FSplice tools have predicted that rs752676953 forms a mutated region that may not be recognized as splice site by the splice site machinery and results in exon skipping which lead to alerted protein function.

Identification of pathogenic missense variants from Cyclin D1 SNPs pool

Out of 169 missense variants, 78 were predicted to be deleterious (Supplementary Table S2). These 79 were further sorted out on basis of their MAF range (0.0001–0.05) and only 4 SNPs (rs557545630, rs534553548. rs535957987, rs143479406) were shortlisted (Table 2) and computationally evaluated for their impact on protein structure and function using different *in silico* tools.

Table 2 Highly deleterious missense substitution having MAF range 0.0001–0.5

RS ID	Missense substitutions	PolyPhen-2*		SIFT†		PROVEAN‡		SNP&GO§		PANTHER‖	
		Scores	Prediction	Scores (0-1)	Prediction	Score	Prediction	Probability	Prediction	Probability	Prediction
rs557545630	NM_053056.2:c.43C>A, NP_444284.1:p.Arg15Ser	0.994	Probably Damaging	0.01	Damaging	-3.8	Deleterious	0.957	Disease	0.361	Neutral
rs534553548	NM_053056.2:c.568G>T, NP_444284.1:p.Ala190Ser	0.894	Possibly damaging	0.12	Tolerant	-2.2	Neutral	0.942	Disease	0.635	Disease
rs535957987	NM_053056.2:c.876C>G, NP_444284.1:p.Asp292Glu	0.994	Probably Damaging	0.15	Tolerant	-2.56	Deleterious	0.785	Disease	0.257	Neutral
rs143479406	NM_053056.2:c.536G>A, NP_444284.1:p.Arg179His,	0.919	Possibly damaging	0.05	Tolerant	-3.84	Deleterious	0.907	Disease	NA	Unclassified
	NM_053056.2:c.536G>T, NP_444284.1:p.Arg179Leu	0.947	Possibly damaging	0.02	Damaging	-5.10	Deleterious	0.837	Disease	0.268	Neutral

*PolyPhen-2 = Polymorphism Phenotyping v2, Scores near to 1 are more confidently predicting the SNP to be damaging. http://genetics.bwh.harvard.edu/pph2/
†SIFT = Sorting Intolerant From Tolerant, SNP scores near 0.00 are more predicted as damaging. http://sift.bii.a-star.edu.sg/
‡PROVEAN = Protein Variation Effect Analyzer, cut off -2.5, scores equal to or above this threshold are predicting the SNP as deleterious http://provean.jcvi.org/index.php
§SNP&GO = scores equal or above 0.5 are predicting the SNP as diseased http://snps.biofold.org/snps-and-go/snps-and-go.html
‖PANTHER = PANTHER scores are given along with SNP&GO scores, scores equal or above 0.5 are predicting the SNP as diseased http://snps.biofold.org/snps-and-go/snps-and-go.html

Table 3 MutPred analysis and prediction of diseased phenotype and molecular mechanism of SNPs

SNP rs ID	Amino acid change	Molecular mechanism	MutPred2 score/ *g* scores	*P*-value	Interpretation*
rs557545630	Arg15Ser	Altered metal binding	0.744	0.03	Disease associated actionable hypotheses
rs534553548	Ala190Ser	NP	0.399		Neutral
rs535957987	Asp292Glu	Loss of relative solvent accessibility	0.753	7.2e-03	Disease associated
		Gain of Strand		0.04	Confident hypotheses
		Altered metal binding		0.01	Confident hypotheses
rs143479406	Arg179His,	NP	0.468	NP	Neutral
	Arg179Leu	NP	0.652	NP	Disease associated

Threshold *P* value $\leq$ 0.05, NP = not predicted
*Interpretation is done on basis of *g* and *P* score; actionable hypotheses: $g > 0.5$, $P<0.05$; confident hypotheses: $g > 0.75$, $P<0.05$ and very confident hypotheses: $g > 0.75$, $P<0.01$. Disease associated when $g > 0.5$

Disease phenotype and molecular mechanism associated with SNPs

MutPred was used to analyse the four selected missense SNPs for prediction of their probability of damaging the protein and molecular mechanism that they can alter (Table 3). It was found that R15S, D292E, and R179L were damaging for protein structure as their *g* score were greater than 0.5. Furthermore, the R15S and D292E were predicted to undergo the molecular changes. R15S has altered the metal binding pocket of Cyclin D1, while D292E has more damaging action that is causing the loss of relative solvent accessibility, the gain of strand and also has altered the metal binding property. The results from FATHMM analysis has shown a low-confidence oncogenic predictions of SNP R15S (P=0.523777), A190S (P=0.740342), R179H (P=0.875302), and R179L (P=0.859033) and association with cancer because the *P* values above 0.5 are predicted to be deleterious. However, D292E (P=0.488885) was predicted to be low-confidence neutral and it may not be associated with cancer. Consistently, the web server SNPeffect results (Supplementary Table S3) has predicted that A190S decreases the aggregation tendency of Cyclin D1 and this SNP results in a ΔG of 1.15 kcal/mol which implies that the mutation has reduces the protein stability. While other SNPs have no effect on aggregation tendency, amyloid propensity and chaperone binding tendency of Cyclin D1. We have also included the SNAP2 prediction results to evaluate the most disease causing SNP. The SNAP2 has predicted R15S and D292 as 'effect' causing SNPs and their scores demonstrates their 'medium' effect for causing disease phenotype.

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

Table 4 PredictSNP analysis results for diseased phenotype

SNP rs ID and amino acid change	Tools and their prediction						Possible impact*
	PredictSNP	CADD	DANN	FATHMM	FunSeq2	GWAVA	
rs557545630 R15S	Neutral	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Diseased
rs534553548 A190S	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Diseased
rs535957987 D292E	Neutral	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral
rs143479406 R179H	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Diseased
rs143479406 R179H	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Diseased

*SNPs predicted as damaging by four or more tools are classified as diseased.

Table 5 Prediction of protein stability using IMutant v.2.0

SNP rs ID	Amino acid change	Stability	RI (0-10)*	DDG (Kcal/mol)*
rs557545630	R15S	Decreased	9	-2.51
rs534553548	A190S	Decreased	9	-0.40
rs535957987	D292E	Increase	3	0.29
rs143479406	R179H	Decreased	8	-1.28
	R179L	Decreased	4	-0.45

*IMutant analysis was accessed at temperature 25°C and pH 7. The reliability index (RI) range is 0–10, where 10 is highest index number, DDG is free energy change value. DDG<0: Decrease Stability and DDG>0: Increase Stability

In order to further validate the effect of SNPs, the PredictSNP analysis was performed and its results displayed that all SNPs are leading towards the disease expect D292E as described in Table 4.

Interestingly, the overall predictions from MutPred, FTHMM, SNP effect, and SNAP2 are not in match with each other; this may be due to low confidence score of SNP in favour of its effect as diseased. However, according to these results and PredictSNP prediction, R15S and A190S may be predicted to cause a more diseased effect with change in the molecular mechanism.

High risk changes in protein stability and evolutionary conservation

The amino acid that are most critical to the protein structure and stability are found to be evolutionary conserved. In order to predict the stability of protein, I-Mutant was accessed. I-Mutant has predicted that all the SNPs caused decreased protein stability except one SNP that is D292E (Table 5). The highest instability found was −0.45 kcal/mol energy related to R179L SNP. Furthermore, HOPE was run to analyse the impact of four SNPs on conservation and protein structure. HOPE has predicted two SNPs (A190S and R179H) that might be highly damaging to the Cyclin D1 protein Table 6. These SNPs can impact the incorrect new bonding and may results in alternation of Cyclin D1 function. The results from SNPeffect has also predicted the instability of Cyclin D1 due to A190S SNP.

The evolutionary conservation of Cyclin D1 amino acid predicted by ConSurf is shown in Figure 3. The SNP position in WT sequence is shown with red-outlined boxes. The amino acid D292 is predicted to be exposed and highly conserved, i.e. a functional residue. Residue R15 and R179 predicted to be exposed while A190 is buried residue. Residue R179 and A190 are also found to be conserved. Any SNP occur in this region may cause damage to the stability and structure of Cyclin D1. We have also performed MSA of human Cyclin D1 with 20 different organisms using Clustal Omega in order to validate the conservation of these residues, the MSA id given in Supplementary File S1, which predicts that R15, R179, A190, and D292 amino acids are highly conserved, any change in amino acid may have a damaging effect on protein structure and its function. Thus, it can be inferred that these missense mutations are conserved in nature, any alternation on their position can lead to unstable protein structure.

Prediction of protein–protein interaction

The Figure 4 and Table 7 indicate the interaction evidence and molecular action of *CCND1* protein (Cyclin D1) with

Table 6 Prediction of change in structure, domain and conservation of Cyclin D1 du to missense SNP using HOPE *in silico* analysis

RS ID	Schematic structure	Amino acid properties	Structure	Domain (InterPro)	Conservation
rs557545630 R15S		The mutant residue is smaller than the WT residue, and might lead to loss of interactions. The WT residue charge was POSITIVE; the mutant residue charge is NEUTRAL, and can cause loss of interactions with other molecules or residues. The mutant residue is more hydrophobic than the WT residue, and can result in loss of hydrogen bonds and/or disturb correct folding.	No impact on structure of protein predicted	No impact on domain was predicted	The mutant residue is located near a highly conserved position and have some properties in common with WT mutated residue. This means that in some rare cases this mutation might occur without damaging the protein.
rs534553548 A190S		The mutant residue is bigger than the WT residue and probably will not fit in core of protein. The hydrophobicity of the WT and mutant residue differs. The mutation will cause loss of hydrophobic interactions in the core of the protein.	In the 3D-structure, it can be seen that the WT residue is located in an α-helix. The mutation converts the WT residue in a residue that does not prefer α-helices as secondary structure.	The residue is buried in the core of a domain. The differences between the WT and mutant residue might disturb the core structure of this domain.	Mutant residue is located near a highly conserved position. This mutation might occur in some rare cases, but it's more likely that the mutation is damaging to the protein.
rs535957987 D292E		The WT residue charge was NEGATIVE, the mutant residue charge is NEUTRAL, which may cause loss of interactions with other molecules or residues. The mutant residue is bigger, this might lead to bumps.	No impact on structure of protein predicted.	No impact on domain was predicted.	The mutant residue was not among the other residue types observed at this position in other, homologous proteins. However, residues that have some properties in common with mutated residue were observed. This means that in some rare cases this mutation might occur without damaging the protein.
rs143479406 R179H		The WT residue charge was POSITIVE; the mutant residue charge is NEUTRAL. Which may cause loss of interactions with other molecules or residues. The mutant residue is smaller than the WT residue. This will cause a possible loss of external interactions.	The WT residue forms a hydrogen bond with glutamic acid at position 162 and Glutamine at position 176, and salt bridge with glutamic acid at position 162, glutamic acid at position 172. The size difference will not makes that the new residue in the correct position to make the same hydrogen bond and salt bridges.	The mutated residue is located on the surface of a domain with unknown function. The residue was not found to be in contact with other domains of which the function is known within the used structure. However, contact with other molecules or domains is still possible and might be affected by this mutation.	The mutant residue is located near a highly conserved position. Neither this mutant residue nor another residue type with similar properties was observed at this position in other homologous sequences. Based on conservation scores this mutation is probably damaging to the protein.

other proteins using STRING. These proteins are meant to be jointly contribute to a shared function with Cyclin D1. The *CCND1* protein is found in strong network signalling with CDK1/2/4/6, CDKN1A/B and in indirect association with MCM10, ORC4, and IFNAR1. In molecular interaction analysis Cyclin D1 is predicted to be directly involve in binding, reaction, catalysis, and inhibition. It is also involving in post-translational modification in association with CDK4/6. Out of these proteins, the CDK4 is a regulatory component of the DC complex which is a major integrator of various mitogenic and antimitogenic signals, and plays a vital role in cancer. The STRING analysis shows that Cyclin D1 is in strong association with the proteins that are playing role in controlling the cell cycle progression and DNA replication events. And also it can be claimed that any change in Cyclin D1 interaction with these proteins can cause change in its associated pathway which can lead to onset of cancer.

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

Table 7 Prediction of molecular interaction of Cyclin D1 with other proteins using STRING

Predicted functional partners	Confidence scores*	Prediction for specific action
CDK4 Cyclin-dependent kinase 4	0.995	Binding, activation, catalysis, reaction, inhibition, expression with inhibition, and post-translational modification
CDK6 Cyclin-dependent kinase 6	0.880	Binding, post-translational modification, catalysis, reaction, and inhibition
CDKN1A Cyclin-dependent kinase inhibitor 1	0.993	Binding, activation, catalysis, reaction, and inhibition
IFNAR1 Interferon alpha/beta receptor 1	0.993	No direct predicted action
MCM10 Protein MCM10 homolog	0.872	No direct predicted action
CDK2 Cyclin-dependent kinase 2	0.853	Binding, activation, catalysis, reaction, and inhibition
CDKN1B Cyclin-dependent kinase inhibitor 1B	0.689	Binding, inhibition catalysis, and reaction
CDKN1C Cyclin-dependent kinase inhibitor 1C	0.814	Binding, reaction, and inhibition
CDK1 Cyclin-dependent kinase 1	0.781	Binding, activation, catalysis, reaction, and inhibition
ORC4 Origin recognition complex subunit 4	0.722	No direct predicted action

*Confidence score = how likely STRING judges an interaction to be true. 0–1, where 1 is highest confidence and ≤ 0.5 represents a false positive interaction.

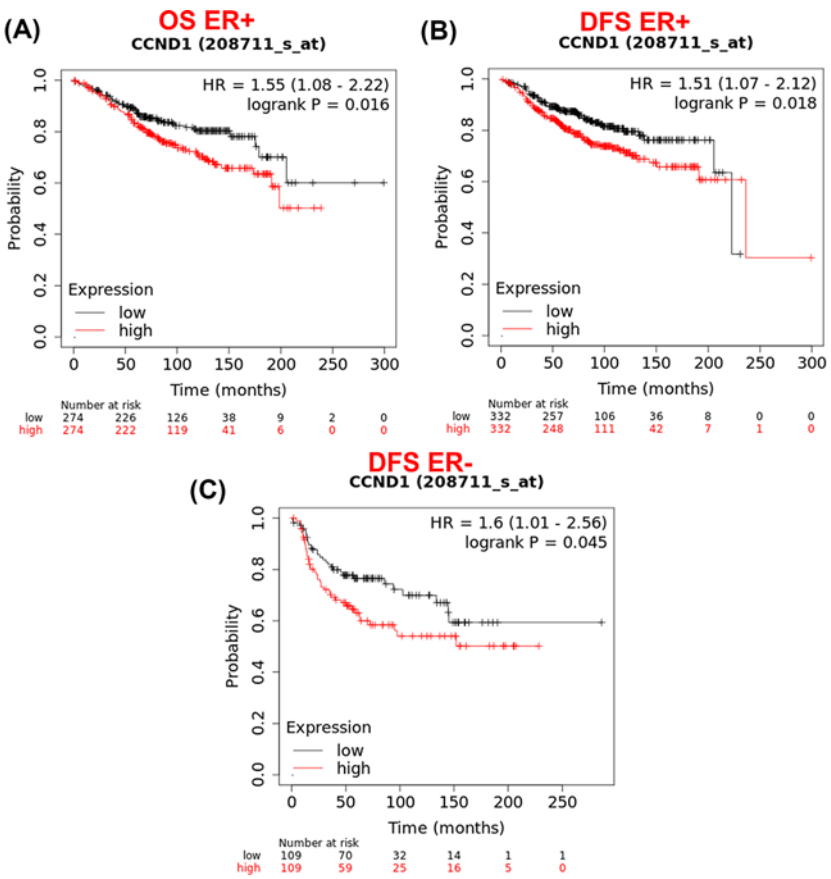


Figure 14. The graphs displaying the positive association of Cyclin D1 expression in patients with breast cancer with different clinical parameters evaluated by Kaplan–Meier plotter

The OS of patients with breast cancer association with (**A**) ER+ (*n*=548), and DFS of patients with breast cancer association with (**B**) ER+ (*n*=664) and (**C**) ER- (*n*=218). On the basis of selected parameters, the analysis was run on '*n*' number of patients with available clinical data; DFS, disease-free survival; OS, overall survival.

Damaging effect of SNP on post-translational modifications sites

The effect of deleterious SNPs on having putative phosphorylation, methylation, sumoylation, and ubiquitination sites was evaluated for change in post-translation modification of Cyclin D1. GPS 5.0 has predicted the significant gain of serine phosphorylation site at R15 and A190 when mutated (R15S and A190S) which was also confirmed by Netphos 3.1. These sites on mutation R15S and A190S are predicted to cause Cyclin D1 prone to Protein Kinase G and Protein kinase phosphorylation, respectively.

In WT Cyclin D1, the GPS-MSP has predicted the R15 and R179 as methylation sites and change in amino acid (R15S and R179H/L) at these sites has predicted to cause loss of methylation function. ModPred has only predicted R179 (score 0.58) as methylation site and upon mutation the loss of methylation will take place. GPS-SUMO and UbPred has not predicted any WT SNP site as sumoylation before or after mutation. The sumoylation and ubiquitination sites were also not predicted by ModPred.

The overall impact of these missense SNPs on post-translational modification of Cyclin D1 can be stated as R15S and A190S can cause a gain of serine phosphorylation sites and loss of methylation function which can lead to cancer progression. While other SNPs has shown no effect on post-translational modification of Cyclin D1.

SNPs impact on Cyclin D1 domain

The WT structure of Cyclin D1 is predominantly consisted of Cyclin domain. Thus, we employed SMART analysis which recognized the Cyclin and Cyclin C domains consisting of 62-146 and 155 to 287 amino acids, respectively, in WT protein structure. These domains were found unaffected by all the four SNPs. However, a non-significant change in *e*-value was noticed due to A190S SNP. ScanProsite detected the Cyclin domain ranging 57-88 amino acid and no effect of SNP was found in this domain. InterPro has shown cyclone-like (IPR013763), Cyclin-N (IPR006671) and Cyclin-C terminal (IPR004367) domain in WT Cyclin D1. These domains were found unaffected by the A15S and D292E SNPs. However, due to R179H, R179L, and A190S SNPs the missing binding sites from Cyclin-like domain (56–131 amino acid) was predicted that is shown in Figure 5. These binding sites are located at the Cyclin box fold which is an integral part of protein binding domain functioning in cell-cycle and transcription control, present in Cyclins, TFIIB and RB. These missing binding sites are positioned at 108, 112, and 119 corresponding to the F, K and L residue respectively. So this can be interpreted that these SNPs might somehow effect the regulation of CDKs. Although these SNPs were not found to effect the domain of Cyclin D1 dominantly; however, A190S may somehow results in change in domain binding site modification.

SNP impact on secondary and tertiary structure of Cyclin D1

Approximately 80% of disease causing variants in amino acid sequence are found in the secondary structure of a protein [16]. Therefore, the effect of SNP on secondary structure of a protein is highly needed to understand any change in tertiary structure of that protein. In this research study, the WT secondary structure analysis by PSIPRED has predicted the helical structure of Cyclin D1 at position R15, R179 and A190, while coiled at D292 as shown in Figure 6. The alternations at these positions has not shown any change in secondary structure conformation except one SNP that is R15S which has caused a secondary structure change in nearby residues as well, that has been shown in sequence plot of PSIPRED in Figure 7. Due to this SNP, the L5 and L6 have been changed from coiled to extracellular conformation; R14 has been changed from coil to helix structure; and A16 from helix to coiled structure change has been predicted. On contradiction, the Phyre.2 analyses has predicted that R15S SNP has caused a loss of alpha helix from position I13, R14, R15 and A16, and this change is also predicted as disordered to native Cyclin D1. The D292E SNP has not caused any change in secondary structure but predicted to have a disordered effect (Figure 8). According to Missense3D prediction, A190S SNP has caused a contraction of cavity volume by 24.192 ^{3}, while R179H SNP has break a salt bridge. The WT salt bridge between NH1 atom of R179 and OE1 atom of E172 has been altered to salt bridge between ND1 atom of H179 and OE1 atom of E162 (distance: 4.595) by A179H SNP. However, overall no structural damage of Cyclin D1 by these SNPs was predicted by Missense3D analysis.

The homology modelling was performed using Swiss-modelling and CPH models. The tertiary structure of Cyclin D1 predicted by Swiss-modelling was from 1 to 265 amino acid (Figure 9), and CPH has predicted the model from 26 to 265 amino acid (Supplementary Figure S1). The structural models of WT Cyclin D1 and mutant predicted by Swiss-modelling are shown in Figure 9. The Swiss-modelling has shown the missing beta sheets due to R15S SNP. This change in Cyclin D1 structure may cause a disruption in protein–protein interaction and other functions [17]. A slight change in protein structure can also be visualized due to A190S SNP, which has caused a contraction in cavity as shown in Figure 9B, where the structure has been visualized from two different directions. These results also confirm the prediction of Phyre.2. For further confirmation of structural predictions, I-Tasser was accessed. The secondary structure results predicted from I-Tasser were also verifying the predictions of Phyre.2. The SNP R15S has cause a

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

loss of helix and gain of coil structure. However, the R179H/L and A190S were also predicted to cause a loss of helix structure effecting amino acid from 12 to 15. I-Tasser has also predicted that these SNPs has not caused any change in solvent accessibility of WT structure. The structural predictions from I-Tasser tool has shown a loss of helix structure around R15 position, due to R15S, R179H/L, and A190S SNPs (Figure 10).

The homology modelling and structural analysis has not predicted a very major change in structure of Cyclin D1 due to these SNPs; however, comparatively R15S and A190S may impact the Cyclin D1 conformation more. Furthermore, as predicted earlier by other *in silico* tools, four SNPs can effect stability, post-translational modification events, and evolutionary conservation of Cyclin D1.

Frameshift mutations

Out of total four frameshift SNPs of Cyclin D1, two were found to be in MAF range (0.0001–0.5) and were selected for structural and functional impact on protein. These two SNPs are rs1565225330 (p.Q176fs) and rs1448866519 (p.L229fs).

MutationTaster has predicted these both SNPs as disease causing that has significantly changed the amino acid sequence and has led to truncated protein. Both of these mutations were also predicted to cause a change in splice site region. Q176fs has cause a gain of donor splice site AT (gDNA position 2852), while L229fs has caused a gain of accepter and donor sites at different splice site regions of Cyclin D1. L229fs was also predicted to cause an existing acceptor site stronger at gDNA position 7016. MutationTaster also predicts the regulatory feature effected by SNPs. These regulatory feature determined a protein to bind to specific DNA sequences and control to switch on genes under any particular conditions. These frameshift SNPs were predicted to possibly affect the histone modification and RNA polymerase feature of Cyclin D1. Q176fs has predicted to cause a change in histone 3 lysine 9 acetylation, histone 3 lysine 4 tri-methylation, histone 3 lysine 36 tri-methylation, histone 3 lysine 4 di-methylation, histone 3 lysine 27 acetylation, histone 3 lysine 79 di-methylation, and RNA polymerase II binding, while L229fs has cause destruction of histone 3 lysine 36 tri-methylation and RNA polymerase II binding. Previously, it has been reported that histone acetylation and histone methylation may inhibit human D1 transcription [18]; therefore, this change in Cyclin D1 regulatory features is predicted to impact it functional activities. MutationTaster uses values from phastCons and phyloP to determine the grade of conservation of a given nucleotide. According to its evolutionary conservation prediction, both the SNPs have caused an alternation at very highly conserved nucleotide.

The tertiary structure of mutant protein was modelled to predicted the structural impact of these two frameshift SNPs on Cyclin D1 using Swiss-modelling and change in domain was predicted by SMART. Figure 11 shows that frameshift mutation has led to formation of the truncated protein and has also lead to loss or formation of damaged Cyclin_C domain due to frameshift SNPs. The drastic change in structure of Cyclin D1 due to these SNPs will cause a loss of Cyclin D1 native function. Q176fs has caused a loss of Cyclin C domain and formed a low compositional complexity region which do not have ability to control the progression of cells through the cell cycle by activating CDK enzymes. On the other hand, L229fs has led to formation of shortened Cyclin C domain which may cause a malfunctioning of Cyclin D1.

Using the SIFT analysis the possibility of damaging effect of frameshift mutation and occurrence of NMD was verified. The outcome from SIFT predicted that both of the SNPs has damaging effect. It has been reported that NMD do not occur when the premature termination codon is in the last exon or it is in the last 50 nucleotides in the second to last exon [19] as shown in Figure 12. Figure 12 is displaying the five exons of the Cyclin D1 and location of possible occurrence of NMD. The position of frameshift SNPs is also mentioned at exon 2 3 and exon 4 of Cyclin D1. The mRNA transcripts that contains the stop codons are eliminated or degraded due to the initiation of NMD which results in the limit of translation of abnormal protein. Q179fs, therefore, may cause NMD and reduce formation of truncated protein.

The structural and functional prediction of frameshift SNPs describes that these two SNPs are forming a truncated Cyclin D1 which is possibly effecting its native function.

UTR SNP effect on Cyclin D1

The UTR SNPs were selected on basis of MAF range. The 116 SNPs of 3′ and 11 SNPs of 5′ UTR SNPs were analyzed using Variowatch, PolymiRTS, and RegulomDB server. The RegulomDB analysis shows that 3′ and 5′ UTR SNPs have not any strong impact to change the regulatory function of *CCND1*. The results also showed that 53/116 SNPs of 3′ UTR and 1/11 SNP of 5′ UTR are less likely to be functional and effect biding (categories 2a, 2b, and 3a) and 63/116 SNPs of 3′ UTR and 10/11 SNPs of 5′ UTR has minimal functional evidence (categories 4, 5, and 6). According to

Table 8 Association of *CCND1* SNPs reported by GWAS

SNPs	Consequence	*P*-values*	Study accession	Studies
rs614367 g.69513996C>T	Intergenic variant	3×10^{-15}	GCST000678	(Turnbull et al. 2010)
		2×10^{-63}	GCST001937	(Ahsan et al. 2014)
		1×10^{-8}	GCST002346	(Michailidou et al. 2013)
rs75915166 g.69564393C>A	Tf-binding site variant	4×10^{-95}	GCST004988	(Michailidou et al. 2015)
		1×10^{-57}	GCST004950	(Michailidou et al. 2017)
rs554219 g.69516874C>A	Regulatory region variant	6×10^{-47}	GCST004988	(Michailidou et al. 2015)
		2×10^{-81}	GCST004950	(Michailidou et al. 2017)
rs78540526 g.69516650C>T	Regulatory region variant	2×10^{-131}	GCST004988	(Michailidou et al. 2015)
		2×10^{-86}	GCST004950	(Michailidou et al. 2017)
		2×10^{-62}	GCST007236	(Michailidou et al. 2015)
rs34507830 g.69646918C>T	Intron variant	7×10^{-31}	GCST004988	(Michailidou et al. 2017)

*Significant *P* value $<10^{-5}$

Table 9 Cox multivariate analysis of prognostic factor association with *CCND1* expression

	OS			DFS		
	HR	**95% CI**	**Log rank***	**HR**	**95% CI**	**Log rank***
ER+	1.55	1.08–2.22	0.016	1.51	1.07–2.12	0.018
ER-	1.54	0.97–2.44	0.067	1.6	1.01–2.56	0.045
HER2+	1.28	0.63–2.62	0.49	0.89	0.47–1.68	0.72
HER2-	2.27	0.88–5.89	0.083	1.32	0.56–3.13	053
PR+	1.4	0.37–5.29	0.62	1.07	0.47–2.42	0.88
PR-	1.22	0.48–3.08	0.68	1.1	0.62–1.97	0.74
Grade 3	1.25	0.9–1.74	0.18	1.25	0.88–1.77	0.21
Lymph node status	1.16	0.78–1.72	0.46	0.81	0.55–1.19	0.27

*Significant $P<0.05$
Abbreviations: DFS, disease-free survival; HR, hazard ratio; OS, overall survival.

the Variowatch and PolymiRTS database prediction, the 3′ and 5′ UTR SNPs have no impact on mRNA stability, interaction (protein–mRNA and miRNA–mRNA) or any influence in *CCND1* expression level. The overall prediction analysis shows that *CCND1* UTR SNPs have no major impact on translation.

Clinical correlation between *CCND1* SNP and patients with breast cancer

CCND1 SNPs were investigated at NCBI to sort out the clinical significant SNPs. Only one SNPs (rs9344) have been reported to have clinical significance. rs9344 is a synonymous SNP and has been associated as a risk factor SNP for different types of cancer [20–25] in different population including breast cancer [22,26]. In Chinese population, this SNP did not show to have a major association with breast cancer. But it has been evaluated that this SNP has an effect of estrogen on breast cancer growth and after diagnosis it predict the survival of patients with breast cancer [27]. Mining the data in GDC data portal, it was found that only one SNP (rs755986542, R260C) was entitled in TCGA data and has shown an in-significant correlation with breast cancer clinically. In a project cases tested for Simple Somatic Mutation (SSM), out of 986 clinical breast cancer cases only 1 case (0.10%) was found as carrier of this SNP, suffering from ductal and lobular neoplasms of stage I.

There are thousands of GWAS studies that associate variants with traits. The most accurate and significant findings (*P*-value $<10^{-5}$) are accomplished by the summary of statistics provided by the association study [28]. In this research study, the comprehensive search was performed to collect all the SNPs of *CCND1* that are associated with breast cancer by GWAS. There are total five *CCND1* SNPs (rs614367, rs75915166, rs554219, rs78540526, and rs34507830) that have shown an association with breast carcinoma in five different GWAS studies as shown in Table 8 . It was found

Bioscience Reports (2021) **41** BSR20202269
https://doi.org/10.1042/BSR20202269

that these SNPs belongs to noncoding region, regulatory region, and transcription binding site region. According to the GWAS, it was also found that among different cancers *CCND1* has highest association with breast cancer.

In the present study, the impact of *CCND1* on survival of patients with breast cancer was investigated using Kaplan–Meier plotter. In Figure 13, the red lines indicate the survival time of patients with breast cancer with high *CCND1* expression levels, and black lines indicate the survival time of patients with breast cancer with low *CCND1* expression levels. Low expression of *CCND1* (0.1761) was found to be correlated with better overall survival (OS) for patients with breast cancer (n=1402). A significant difference was observed between OS and disease-free survival (DFS) of patients with breast cancer (n=1746), which indicates that the patients with *CCND1* alternations has improved prognosis as compared with those without *CCND1* alternations. However, a strong difference in curve is noted between low and high expression level, which shows that the high expression of *CCND1* is found to be associated with high number of patients at risk which gives a less survival rate for patients with breast cancer.

To further investigate the association of Cyclin D1 expression in breast cancer patients with different clinical parameters, HER2- (P=0.083), ER- (P=0.016), and ER+ (P=0.067) were found significant associated with OS and ER+ (P=0.018) and ER- (P=0.045) were significant with DFM in breast cancer (Figure 14). While there was no statistical evidence which reveals the association of PR and HER2 with DFM with expression of Cyclin D1 in patients with breast cancer (Supplementary Figure S2). The prognostic factors association with *CCND1* expression is also shown by cox multivariate analysis given in Table 9.

The variants with unknown clinical significance are classified by the American College of Medical Genetics as Variants of Uncertain Significance (VUS). The *in silico* tools, like SIFT, PolyPhen-2 and FATHMM, have been used previously to predict the impact of VUS on pathogenicity of its mutant protein, while their use to predict the clinical significance of a particular SNP or mutation is still unclear [29]. In this research study, on basis of analysis from these tools, out of 169 missense SNPs of *CCND1* the 77 have been predicted as highly deleterious SNPs and have been associated with a diseased phenotype (Supplementary Table S2). To further significantly validate the SNP with its clinical importance, an extensive *in vitro* and *in vivo* research study is needed.

Discussion

Breast cancer is worldwide emerging disorder of women with approximately 2.1 million new cases are diagnosed and almost 0.6 million deaths in each year [30]. The overall survival rate of patients with breast cancer has reached up to 90%,;however, metastatic or advanced breast cancer survival rate is still 25% [31]. Recent therapeutic approach which includes endocrine therapy and targeted therapy is productive in prognosis of breast cancer. Furthermore, Cyclin-dependent kinase such as CDK4 and CDK6 inhibitors also showed clinical benefits [32]. In the present study, we have collected *CCND1* SNPs data from the NCBI genome workbench and applied *in silico* analysis on coding nonsynonymous SNPs, splice site SNPs and 5′ and 3′ UTR SNP to predict their pathogenic impact on protein structure and function in relation with breast cancer.

Our results displayed that four SNPs reside in highly conserved region of Cyclin D1. However, only R15S and A190S have displayed a significant diseased phenotype and an altered molecular mechanism predicted by MutPred, FATHMM, SNPeffect, SNAP2, and PredictSNP. Further analysis indicated that A190S, R179L, and R15S may also cause a decrease in stability of Cyclin D1 anticipated by I-Mutant, HOPE, and SNPeffect. Previous study on the mutation in splicing region showed that the improper exon and intron recognition occurred by splicing machinery which results in an aberrant transcript. These variants may also interrupt the existing splice sites, create new ones and also it can impact splicing enhancers and silencers binding. Therefore, the variants in splice site accounted for most of the diseases [15]. Generally, the variants residing within 10 nucleotide position in intronic splice site region account more in creating defective splice site [15]. The two most occurring splice transcripts of *CCND1* are Cyclin D1a and Cyclin D1b. Cyclin D1a contains all the 5 exons but Cyclin D1b formed by intron 4 inclusion and exon 5 skipping which is modulated by Serine and arginine rich splicing factor 1 (SRSF1) for up-regulation of Cyclin D1b in breast cancer. The Cyclin D1b is a product of alternate splicing due to silent polymorphism G/A870, in which the A allele was assumed to reduce the efficacy of the splice donor site [33]. In our study, 23 SNPs out of total 2506 intronic SNPs were found within 10 nucleotide position at 5′ and 3′ region of intron. The SNP declared as damaging by three or more *in silico* tools out of five (HSF, Sroogle, SliceView, NetGene2, and FSplice) is considered as highly pathogenic and may effects the splicing region. The *in silico* analysis of these tools has predicted that the SNP can affect the donor or acceptor site of splice region of *CCND1*.

In molecular interaction analysis, Cyclin D1 is predicted to be directly involve in binding, reaction, catalysis, and inhibition. It is also take part in post-translational modification in association with CDK4/6. The CDK plays a vital role in controlling the G1/2-S transition in cell cycle, promotes the E2F transcriptional program, and the initiation

of DNA synthesis. These are also found to be involved in the assembly, stability, and modulation of DC activation. Out of these proteins, the CDK4 is a regulatory component of the DC complex which is a major integrator of various mitogenic and antimitogenic signals, plays a vital role in cancer [34,35]. We performed the STRING analysis which demonstrated that Cyclin D1 remain in strong association with the proteins that play vital role in cell cycle progression and DNA replication events. Thus, it can be suggested that any change in Cyclin D1 interaction with these proteins may change its associated pathway which can lead to onset of cancer. The WT structure of Cyclin D1 is predominantly consisted of Cyclin domain. The Cyclin domain is generally present in Cyclins proteins, and also transcription factor IIB (TFIIB) and retinoblastoma (RB). These functioning in cell-cycle, transcription control, and cancer progression [36]. In order to understand the role of Cyclin domain of *CCND1*, SMART software was used and analysis revealed that Cyclin and Cyclin C domains consisting of 62–146 and 155–287 amino acids, respectively, in WT protein structure. These domains remain unaffected by our SNP analysis; however, a nonsignificant change in *e*-value was noted. Furthermore, ScanProsite spotted the Cyclin domain range from 57 to 88 amino acid while no effect of SNP was observed in this domain. InterPro has shown cyclone-like (IPR013763), Cyclin-N (IPR006671), and Cyclin-C terminal (IPR004367) domain in WT Cyclin D1. These domains were found unaffected by the A15S and D292E SNPs. What is more, effect of damaging SNP was predicted on secondary structure Cyclin D1 and we observed contraction of cavity volume while homology modelling, and structural analysis has not predicted a very major change in structure of Cyclin D1. Next, we studied survival of patients with breast cancer and prognosis prediction using Kaplan–Meier plotter. We found low expression of *CCND1* (0.1761) was found to be correlated with better OS for patients with breast cancer (n=1402). Further analysis indicated that there is a significant difference between OS and DFS of patients with breast cancer (n=1746), which indicates that the patients with *CCND1* alternations has improved prognosis as compared with those without *CCND1* alternations.

Altogether, out of 3747 SNPs of *CCND1*, only one splice site SNP rs752676953 (c.1988+5G<A) and two frameshift SNPs, rs1565225330 (p.Q176fs) and rs1448866519 (p.L229fs), have predicted to be strongly effect the splice site. Two missense SNPs rs557545630 (R15S) and rs534553548 (A190S) have diseased phenotype which may also affect the post-translational modification of Cyclin D1. A190S was predicted to cause a major change in stability of the Cyclin D1 and may somehow bring changes in domain binding site. No major change in structure of Cyclin D1 by these missense SNPs was predicted; however, two frameshift SNPs have resulted truncated Cyclin D1 protein structure. These changes might affect the Cyclin D1 interaction with other proteins and can disrupt its function as well. None of these SNPs were previously related with breast cancer. However, GWAS study has reported 5 SNPs (rs614367, rs75915166, rs554219, rs78540526, and rs34507830) to be significantly associated with breast cancer. The Kaplan–Meier plotter has explained that high expression of *CCND1* is associated with less survival rate of breast cancer patients. Our study suggests that c.1988+5G<A, R15S, R179L, and A190S SNPs could directly or indirectly destabilize the Cyclin D1. If promise of the computational analysis of SNPs is to be realised, this information can be integrated into *in vivo* and *in vitro* analysis to further validate and implement these SNPs in treatment or prognosis of breast cancer.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

The authors declare that there are no sources of funding to be acknowledged.

Author Contribution

Conceived and designed: A.A., R.K. Performed the Bioinformatics analysis: A.A., A.U., M.A., W.S. A.H. Analysed the data: R.K., A.A., H.M.J.H. Wrote the paper: A.A., R.K. Modification of the manuscript: W.S., A.W., H.M.J.H.

Acknowledgements

This article is part of Ms. Ayesha Aftab (79-FBAS/PhDBT/S17) PhD thesis.

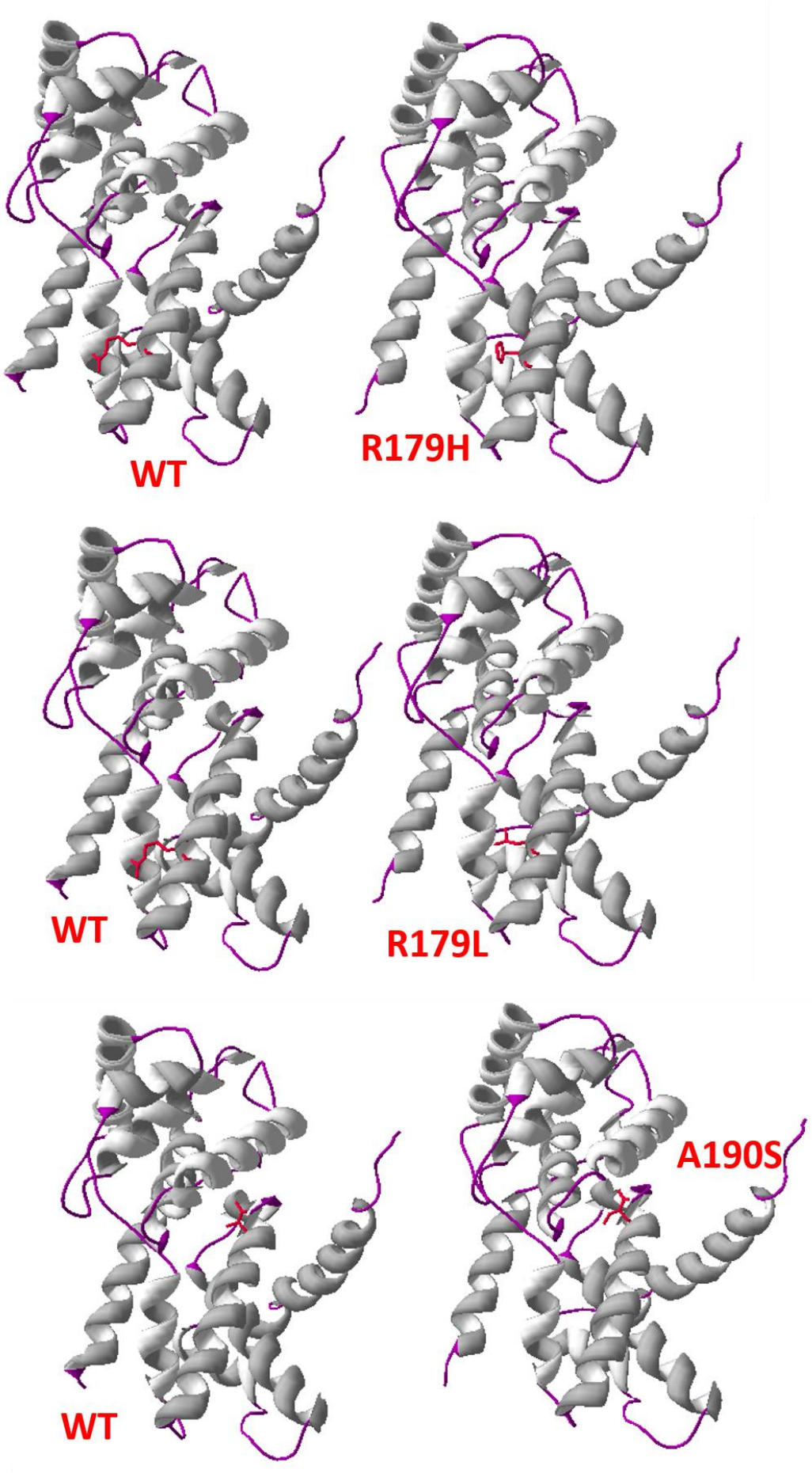
Abbreviations

DFM, disease-free survival; HR, hazard ratio; OS, overall survival; SNP, single-nucleotide polymorphism; SS, splice site.

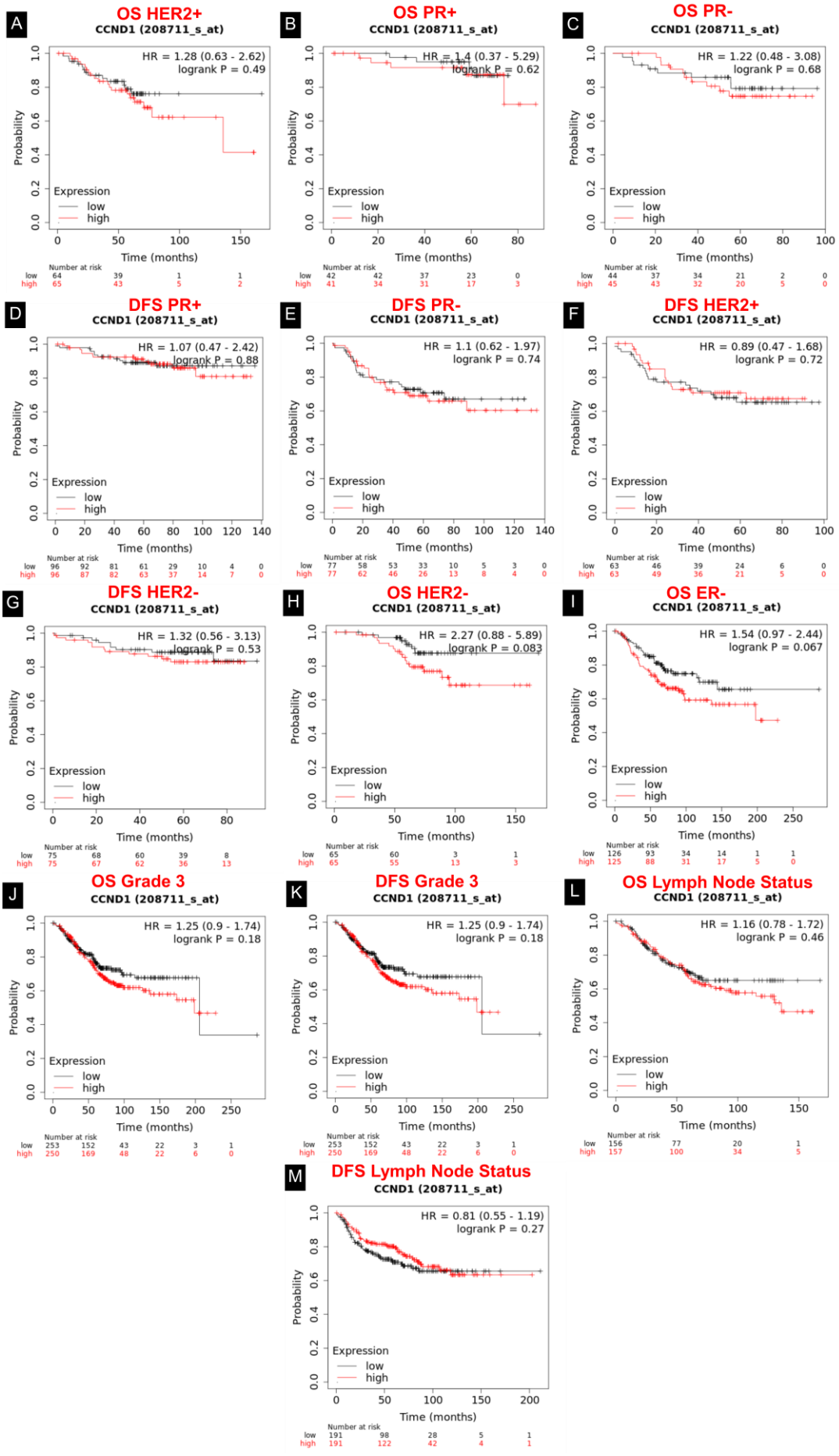
References

1 Akram, M., Iqbal, M., Daniyal, M. and Khan, A.U. (2017) Awareness and current knowledge of breast cancer. *Biol. Res.* **50**, 33–56, https://doi.org/10.1186/s40659-017-0140-9
2 Coleman, M.P., Quaresma, M., Berrino, F. et al. (2008) Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol.* **9**, 730–756, https://doi.org/10.1016/S1470-2045(08)70179-7
3 Reyna, C. and Lee, M.C. (2014) Breast cancer in young women: Special considerations in multidisciplinary care. *J. Multidiscip. Healthc* **7**, 419–429
4 Biéche, I., Olivi, M., Noguès, C., Vidaud, M. and Lidereau, R. (2002) Prognostic value of CCND1 gene status in sporadic breast tumours, as determined by real-time quantitative PCR assays. *Br. J. Cancer* **86**, 580–586, https://doi.org/10.1038/sj.bjc.6600109
5 Jayasurya, R., Sathyan, K.M., Lakshminarayanan, K., Abraham, T., Nalinakumari, K.R., Abraham, E.K. et al. (2005) Phenotypic alterations in Rb pathway have more prognostic influence than p53 pathway proteins in oral carcinoma. *Mod. Pathol.* **18**, 1056–1066, https://doi.org/10.1038/modpathol.3800387
6 Sathyan, K.M., Nalinakumari, K.R., Abraham, T. and Kannan, S. (2008) CCND1 polymorphisms (A870G and C1722G) modulate its protein expression and survival in oral carcinoma. *Oral Oncol.* **44**, 1056–1066, https://doi.org/10.1016/j.oraloncology.2007.09.003
7 Knudsen, K.E., Alan Diehl, J., Haiman, C.A. and Knudsen, E.S. (2006) Cyclin D1: Polymorphism, aberrant splicing and cancer risk. *Oncogene* **25**, 1620–1628, https://doi.org/10.1038/sj.onc.1209371
8 Izzo, J.G., Wu, T.T., Wu, X. et al. (2007) Cyclin D1 guanine/adenine 870 polymorphism with altered protein expression is associated with genomic instability and aggressive clinical biology of esophageal adenocarcinoma. *J. Clin. Oncol.* **25**, 698–707
9 Wang, T.C., Cardiff, R.D., Zukerberg, L., Lees, E., Arnold, A. and Schmidt, E.V. (1994) Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**, 669–671, https://doi.org/10.1038/369669a0
10 Shastry, B.S. (2009) SNPs: impact on gene function and phenotype. *Methods Mol. Biol.* **578**, 3–22, https://doi.org/10.1007/978-1-60327-411-1˙1
11 Capriotti, E., Fariselli, P. and Casadio, R. (2005) I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* W306–W310, https://doi.org/10.1093/nar/gki375
12 Boyle, A.P., Hong, E.L., Hariharan, M. et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797, https://doi.org/10.1101/gr.137323.112
13 Chen, Y.H., Liu, C.K., Chang, S.C., Lin, Y.J., Tsai, M.F., Chen, Y.T. et al. (2008) GenoWatch: a disease gene mining browser for association study. *Nucleic Acids Res.* **36**, W336–W340, https://doi.org/10.1093/nar/gkn214
14 Bao, L., Zhou, M., Wu, L., Lu, L., Goldowitz, D., Williams, R.W. et al. (2007) PolymiRTS Database: Linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res.* **35**, 51–54, https://doi.org/10.1093/nar/gkl797
15 Anna, A. and Monika, G. (2018) Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* **59**, 253–268, https://doi.org/10.1007/s13353-018-0444-7
16 Khan, S. and Vihinen, M. (2007) Spectrum of disease-causing mutations in protein secondary structures. *BMC Struct. Biol.* **7**, https://doi.org/10.1186/1472-6807-7-56
17 Craveur, P., Joseph, A.P., Rebehmed, J. and De Brevern, A.G. (2013) β-Bulges: Extensive structural analyses of β-sheets irregularities. *Protein Sci.* **22**, 1366–1378, https://doi.org/10.1002/pro.2324
18 Guo, Z.Y., Hao, X.H., Tan, F.F., Pei, X., Shang, L.M., Jiang, X.L. et al. (2011) The elements of human cyclin D1 promoter and regulation involved. *Clin. Epigenetics* **2**, 63–76, https://doi.org/10.1007/s13148-010-0018-y
19 Lindeboom, R.G.H., Supek, F. and Lehner, B. (2016) The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118, https://doi.org/10.1038/ng.3664
20 Qiu, H., Cheng, C., Wang, Y., Kang, M., Tang, W., Chen, S. et al. (2016) Investigation of cyclin D1 rs9344 G>A polymorphism in colorectal cancer: A meta-analysis involving 13,642 subjects. *Onco. Targets Ther.* **9**, 6641–6650
21 Hussain, S., Yuvaraj, M., Thakur, N., Salam, I., Singh, N., Mir, M.M. et al. (2011) Association of cyclin D1 gene polymorphisms with risk of esophageal squamous cell carcinoma in Kashmir Valley-A high risk area. *Mol. Carcinog.* **50**, 487–498, https://doi.org/10.1002/mc.20732
22 Thakur, N., Kumari, S. and Mehrotra, R. (2018) Association between Cyclin D1 G870A (rs9344) polymorphism and cancer risk in Indian population: Meta-analysis and trial sequential analysis. *Biosci. Rep.* **38**, BSR20180694, https://doi.org/10.1042/BSR20180694
23 Zhang, Y., Zeng, X., Lu, H., Ji, H., Zhao, E. and Li, Y. (2016) Association between cyclin D1 (CCND1) G870A polymorphism and gastric cancer risk: A meta-analysis. *Oncotarget* **7**, 66109–66118, https://doi.org/10.18632/oncotarget.11848
24 Xie, M., Zhao, F., Zou, X., Jin, S. and Xiong, S. (2017) The association between CCND1 G870A polymorphism and colorectal cancer risk. *Med. (United States)* **96**, e8269
25 Hu, Y.Y., Zheng, R., Guo, C. and Niu, Y.M. (2014) Association between cyclin D1 G870A polymorphism and cervical cancer risk: A cumulative meta-analysis involving 2,864 patients and 3,898 controls. *Diagn. Pathol.* **9**, 168–174, https://doi.org/10.1186/s13000-014-0168-x
26 Yaylim-Eraltan, I., Ergen, A., Görmüş, U., Arikan, S., Küçücük, S., Şahin, O. et al. (2009) Breast cancer and cyclin D1 gene polymorphism in Turkish women. *In Vivo (Brooklyn)* **23**, 767–772
27 Xiao, O.S., Moore, D.B., Cai, Q., Cheng, J., Wen, W., Pierce, L. et al. (2005) Association of cyclin D1 genotype with breast cancer risk and survival. *Cancer Epidemiol. Biomarkers Prev.* **14**, 91–97
28 Bush, W.S. and Moore, J.H. (2012) Chapter 11: Genome-Wide Association Studies. *PLoS Comput. Biol.* **8**, e1002822, https://doi.org/10.1371/journal.pcbi.1002822
29 Kerr, I.D., Cox, H.C., Moyes, K. et al. (2017) Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *J. Community Genet.* **8**, 87–95, https://doi.org/10.1007/s12687-016-0289-x

30 Wu, Y., Zhang, Y., Pi, H. and Sheng, Y. (2020) Current Therapeutic Progress of CDK4/6 Inhibitors in Breast Cancer. *Cancer Manag. Res. Volume* **12**, 3477–3487, https://doi.org/10.2147/CMAR.S250632
31 Serra, F., Lapidari, P., Quaquarini, E., Tagliaferri, B., Sottotetti, F. and Palumbo, R. (2019) Palbociclib in metastatic breast cancer: Current evidence and real-life data. *Drugs Context* **8**, 1–16, https://doi.org/10.7573/dic.212579
32 Baselga, J., Campone, M., Piccart, M. et al. (2012) Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N. Engl. J. Med.* **366**, 520–529, https://doi.org/10.1056/NEJMoa1109653
33 Betticher, D.C., Thatcher, N., Altermatt, H.J., Hoban, P., Ryder, W.D.J. and Heighway, J. (1995) Alternate splicing produces a novel cyclin D1 transcript. *Oncogene* **11**, 1005–1011
34 Chen, K., Jiao, X., Ashton, A. et al. (2020) The membrane-associated form of cyclin D1 enhances cellular invasion. *Oncogenesis* **9**, 83–96, https://doi.org/10.1038/s41389-020-00266-y
35 Hinds, P.W., Dowdy, S.F., Eaton, E.N., Arnold, A. and Weinberg, R.A. (1994) Function of a human cyclin gene as an oncogene. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 709–713, https://doi.org/10.1073/pnas.91.2.709
36 Yang, V.W. (2018) The Cell Cycle. *Physiol. Gastrointest. Tract*, Sixth Ed., pp. 197–219, Elsevier Inc.


WT
R179H
WT
R179L
WT
A190S

Supplementary Figure 1. The CPH model of WT cyclin D1 structure, and mutant cyclin D1 structures due to R179H, R17L and A190S SNPs


A
OS HER2+
CCND1 (208711_s_at)
HR = 1.28 (0.63 - 2.62)
logrank P = 0.49
B
OS PR+
CCND1 (208711_s_at)
HR = 1.4 (0.37 - 5.29)
logrank P = 0.62
C
OS PR-
CCND1 (208711_s_at)
HR = 1.22 (0.48 - 3.08)
logrank P = 0.68
D
DFS PR+
CCND1 (208711_s_at)
HR = 1.07 (0.47 - 2.42)
logrank P = 0.88
E
DFS PR-
CCND1 (208711_s_at)
HR = 1.1 (0.62 - 1.97)
logrank P = 0.74
F
DFS HER2+
CCND1 (208711_s_at)
HR = 0.89 (0.47 - 1.68)
logrank P = 0.72
G
DFS HER2-
CCND1 (208711_s_at)
HR = 1.32 (0.56 - 3.13)
logrank P = 0.53
H
OS HER2-
CCND1 (208711_s_at)
HR = 2.27 (0.88 - 5.89)
logrank P = 0.083
I
OS ER-
CCND1 (208711_s_at)
HR = 1.54 (0.97 - 2.44)
logrank P = 0.067
J
OS Grade 3
CCND1 (208711_s_at)
HR = 1.25 (0.9 - 1.74)
logrank P = 0.18
K
DFS Grade 3
CCND1 (208711_s_at)
HR = 1.25 (0.9 - 1.74)
logrank P = 0.18
L
OS Lymph Node Status
CCND1 (208711_s_at)
HR = 1.16 (0.78 - 1.72)
logrank P = 0.46
M
DFS Lymph Node Status
CCND1 (208711_s_at)
HR = 0.81 (0.55 - 1.19)
logrank P = 0.27
Probability
Expression
low
high
Time (months)
Number at risk

Supplementary Figure 2. The association of cyclin D1 with different clinical parameters in breast cancer patients. OS association with **A)** HER+ (n= 129), **B)** PR+ (n= 83) and **C)** PR- (n= 89). DFM association of breast cancer patients with **D)** PR+ (n= 192), **E)** PR- (n= 154), **F)** HER+ (n= 126), and **G)** HER- (n= 150)

Supplementary Table 1. *In silico* analysis of intronic splice region *CCND1* SNPs.

No.	RS ID	SNP	HSF [a] HSF matrix score (0-100)				SpliceView [b] Score (0-100)		Sroogle [c]			NetGene2 [d] confidence score (0-1)		FSplice [e] (Weight index)		Prediction of possible impact [f]
			WT	Mu	Variation %	Interpretation	WT	Mu	Element	WT	Mu	WT	Mu	WT	Mu	
1	rs374405138	g.5411C>T, c.198+4C>T	44.25	71.08	+60.63	Potential alteration of splicing. New Donor Site	84 DS	84 DS	5´ SS	-3.80 (Delta-G) 6.79 (Max entropy)	-3.90 (Delta-G) 4.48 (Max entropy)	0.70 DS	0.61 DS	7.36	6.66	No Effect
2	rs752676953	g.5412G>A, c.198+5G>A	86.34	74.18	-14.08	Probably effecting splicing	84 DS	78 DS	5´ SS	-3.80 (Delta-G) 6.79 (Max entropy)	-3.80 (Delta-G) 2.02(Max entropy)	0.70 DS	Not generated	W= 7.36	Not generated	Probably effecting
3	rs1456525574	g.5413G>A, c.198+6G>A	86.34	85.94	-0.46	Probably no impact on splicing	84 DS	83 DS	5´ SS	-3.80 (Delta-G) 6.79 (Max entropy)	-3.80 (Delta-G) 5.51 (Max entropy)	0.70 DS	0.71 DS	7.36	W=7.92	No Effect
4	rs1343339113	g.5415G>A, c.198+8G>A	No difference between mutant and reference sequence			Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			0.70 DS	0.52 DS	7.36	7.36	No Effect
5	rs1420743674	g.6919C>T, c.199-8C>T	82.96	82.34	-0.75	Probably no impact on splicing	84 AS	84 AS	3´ SS	9.86 (Max entropy) 80.21 (PSSM)	9.54 (Max entropy) 80.80 (PSSM)	0.96	0.96	4.03	4.03	No Effect
6	rs758963834	g.6922C>T, c.199-5C>T	82.96 AS	83.85 AS	+1.07	Probably no impact on splicing	84 AS	84 AS	3´ SS	9.86 (Max entropy) 80.21 (PSSM)	9.51 (Max entropy) 81.22 (PSSM)	0.96	0.96	4.03	4.03	No Effect
7	rs1339178943	g.6923G>A, c.199-4G>A	82.96 AS	83.03 AS	+0.08	Probably no impact on	84 AS	84 AS	3´ SS	9.86 (Max	10.31 (Max entropy)	0.96 AS	0.97 AS (H)	4.03	4.53	No Effect

						splicing				entropy) 80.21 (PSSM)	80.45 (PSSM)					
8	rs1424359226	g.7147C>G, c.414+5C>G	81.06 DS	93.08 DS	+14.83	Probably no impact on splicing	81 DS	82 DS	5′ SS	-3.60 (Delta-G) 4.51 (Max entropy)	-5.90 (Delta-G) 8.69 (Max entropy)	0.70 DS	0.99 DS (H)	6.24	12.82	No Effect
9	rs1484194988	g.7151delT, c.414+9delT	No difference between mutant and reference sequence			Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			0.70 DS	0.39 DS	6.24	6.24	No Effect
10	rs762325000	g.7720G>A, c.415-8G>A	85.61 AS	85.54 AS	-0.08	Probably no impact on splicing	83 AS	83 AS	3′ SS	7.44 (Max entropy) 81.57 (PSSM)	6.22 (Max entropy) 81.61 (PSSM)	0.34 AS	0.20	8.20	6.83	No Effect
11	rs1240440953	g.7723C>T, c.415-5C>T	85.61 AS	86.50 AS	+1.04	Probably no impact on splicing	83 AS	84 AS	3′ SS	7.44 (Max entropy) 81.57 (PSSM)	8.01 (Max entropy) 82.58 (PSSM)	0.56	0.56	8.20	9.57	No effect
12	rs377200375	g. 7894C>T, c.574+7C>T	56.84 DS	83.68 DS	+47.22	Potential alteration of splicing. New Donor Site	Not predicted	Not predicted	No difference between mutant and reference sequence			0.95 DS (H)	0.91 DS (H) New donor site	11.00	11.00	No effect
13	rs931931114	g.7896delG, c.574+9delG	76.89 AS	72.73 AS	-5.41	Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			0.95 DS (H)	0.95 DS (H)	11.00	11.00	No effect
14	rs750369077	g.7897G>T, c.574+10G>T	76.89 AS	72.59 AS	-5.59	Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			0.95 DS (H)	0.95 DS (H)	11.00	11.00	No effect
15	rs201881393	g.11882T>C, c.575-8T>C	89.65 AS	90.26 AS	+0.68	Probably no impact on splicing	92 AS	92AS	3′ SS	10.04 (Max entropy) 88.71 (PSSM)	9.82 (Max entropy) 88.12 (PSSM)	1.00 AS (H)	1.00 AS (H)	13.18	12.30	No effect
16	rs1443089370	g.11883G>A, c.575-7G>A	89.65 AS	89.69 AS	+0.04	Probably no impact on	92 AS	92 AS	3′ SS	10.04 (Max	9.68 (Max entropy)	1.00 AS	1.00 AS (H)	13.18	12.93	No effect

						splicing				entropy) 88.71 (PSSM)	89.06 (PSSM)	(H)				
17	rs776761881	g.11885C>T, c.575-5C>T	89.65 AS	90.54 AS	+0.99	Probably no impact on splicing	92 AS	92 AS	3´ SS	10.04 (Max entropy) 88.71 (PSSM)	10.79 (Max entropy) 89.72 (PSSM)	1.00 AS (H)	1.00 AS (H)	13.18	13.30	No effect
18	rs1398886316	g.12045G>A, c.723+7G>A	84.35 DS	85.51 DS	+1.38	Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			1.00 DS (H)	1.00 DS (H)	16.18	16.18 6.94 (new DS)	No effect
19	rs1389923822	g.12047G>A, c.723+9G>A	84.35 DS	72.18 DS	-14.43	Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			1.00 DS (H)	1.00 DS (H)	16.18	16.18 7.22 (new DS)	No effect
20	rs371455093	g.12048G>C, c.723+10G>C	84.35 DS	84.11 DS	-0.28	Probably no impact on splicing	Not predicted	Not predicted	No difference between mutant and reference sequence			1.00 DS (H)	1.00 DS (H)	16.18	16.18 5.26 (new DS)	No effect
21	rs753060017	g.15004T>C, c.724-10T>C	93.38 AS	92.67 AS	-0.76	Probably no impact on splicing	Not predicted	Not predicted	3´ SS	10.71 (Max entropy) 88.03 (PSSM)	10.63 (Max entropy) 87.02 (PSSM)	1.00 AS (H)	1.00 AS (H)	13.60 AS	12.97	No effect
22	rs571153521	g.15008T>C, c.724-6T>C	93.38 AS	94.03 AS	+0.7	Probably no impact on splicing	93 AS	93 AS	3´ SS	10.71 (Max entropy) 88.03 (PSSM)	10.98 (Max entropy) 87.93 (PSSM)	1.00 AS (H)	1.00 AS (H)	13.60 AS	10.60	No effect
23	rs764630402	g.15010T>G, c.724-4T>G	93.38 AS	93.41 AS	+0.03	Probably no impact on splicing	93 AS	93 AS	3´ SS	10.71 (Max entropy) 88.03 (PSSM)	11.20 (Max entropy) 87.51 (PSSM)	1.00 AS (H)	1.00 AS (H)	13.60 AS	10.60	No effect

Supplementary Table 2. *In silico* analysis of *CCND1* reported missense SNPs.

	RS ID	Missense substitutions	PolyPhen-2 [a]		SIFT [b]		PROVEAN [c]		SNP & GO [d]		PANTHER [e]		Prediction of possible impact [f]
			scores	Prediction	Scores (0-1)	Prediction	Score	Prediction	Probability	Prediction	Probability	Prediction	
1	rs572037183	NM_053056.2:c.4G>A, NP_444284.1:p.Glu2Gln	0.145	Benign	0.15	Tolerant	-0.84	Neutral	0.581	Disease	NA	Unclassified	Benign
		NM_053056.2:c.4G>C, NP_444284.1:p.Glu2Lys	0.023	Benign	0.13	Tolerant	-0.96	Neutral	0.803	Disease	NA	Unclassified	Benign
2	rs545664320	NM_053056.2:c.8A>G, NP_444284.1:p.His3Arg	0.244	Benign	0.37	Tolerant	-1.3	Neutral	0.665	Disease	NA	Unclassified	Benign
3	rs1356268873	NM_053056.2:c.12G>T, NP_444284.1:p.Gln4His	0.904	Probably Damaging	0.02	Damaging	-1.49	Neutral	0.431	Neutral	NA	Unclassified	Benign
4	rs993495966	NM_053056.2:c.19T>A, NP_444284.1:p.Cys7Ser	0.999	Probably Damaging	0.01	Damaging	-6.75	Deleterious	0.978	Disease	NA	Unclassified	HighlyDamaging
5	rs1486252402	NM_053056.2:c.20G>C, NP_444284.1:p.Cys7Ser	0.999	Probably Damaging	0.01	Damaging	-6.75	Deleterious	0.978	Disease	NA	Unclassified	Highly Damaging
6	rs773884084	NM_053056.2:c.22T>G, NP_444284.1:p.Cys8Gly	0.021	Benign	0.06	Tolerant	-6.46	Deleterious	0.978	Disease	0.421	Neutral	Benign
7	rs1340132260	NM_053056.2:c.23G>A, NP_444284.1:p.Cys8Tyr	0.919	Probably Damaging	0.19	Tolerant	-5.69	Deleterious	0.958	Disease	0.289	Neutral	Highly Damaging
8	rs761266790												
9	rs1372181670	NM_053056.2:c.40C>A, NP_444284.1:p.Arg14Ser	0.411	Benign	0.74	Tolerant	-2.13	Neutral	0.879	Disease	0.282	Neutral	Benign
10	rs557545630	NM_053056.2:c.43C>A, NP_444284.1:p.Arg15Ser	0.994	Probably Damaging	0.01	Damaging	-3.8	Deleterious	0.957	Disease	0.361	Neutral	Highly Damaging
11	rs1299107729	NM_053056.2:c.46G>T, NP_444284.1:p.Ala16Ser	0.998	Probably Damaging	0.06	Tolerant	-1.74	Neutral	0.961	Disease	0.565	Disease	Highly Damaging
12	rs772857967	NM_053056.2:c.52C>G, NP_444284.1:p.Pro18Ala	0.001	Benign	0.75	Tolerant	-2.34	Neutral	0.695	Disease	0.293	Neutral	Benign
		NM_053056.2:c.52C>T, NP_444284.1:p.Pro18Ser	0.000	Benign	0.70	Tolerant	-2.26	Neutral	0.780	Disease	0.266	Neutral	Benign
13	rs1417631865	NM_053056.2:c.57T>A, NP_444284.1:p.Asp19Glu	0.987	Probably Damaging	0.05	Tolerant	-2.85	Deleterious	0.960	Disease	0.444	Neutral	Highly Damaging
14	rs766170770	NM_053056.2:c.61A>G,	0.645	Possibly	0.36	Tolerant	-2	Neutral	0.810	Disease	0.296	Neutral	Benign

		NP_444284.1:p.Asn21Asp		damaging									
15	rs753863475	NM_053056.2:c.67C>T, NP_444284.1:p.Leu23Phe	0.01	Benign	0.06	Tolerant	-2.11	Neutral	0.876	Disease	0.487	Neutral	Benign
16	rs1400812649	NM_053056.2:c.74A>G, NP_444284.1:p.Asp25Gly	0.876	Possibly damaging	0.01	Damaging	-4.51	Deleterious	0.986	Disease	0.800	Disease	Highly Damaging
17	rs2220247	NM_053056.2:c.88G>T, NP_444284.1:p.Ala30Ser,	0.000	Benign	0.47	Tolerant	0.79	Neutral	0.628	Disease Disease	0.187	Neutral	Benign
		NM_053056.2:c.88G>A, NP_444284.1:p.Ala30Thr	0.000	Benign	0.62	Tolerant	0.16	Neutral	0.658	Disease	0.208	Neutral	Benign
18	rs1415272481	NM_053056.2:c.89C>A, NP_444284.1:p.Ala30Asp	0.074	Benign	0.07	Tolerant	-0.43	Neutral	0.933	Disease	0.355	Neutral	Benign
19	rs746088878	NM_053056.2:c.106G>A, NP_444284.1:p.Glu36Lys	0.116	Benign	0.02	Damaging	-2.8	Deleterious	0.803	Disease	0.233	Neutral	Highly Damaging
20	rs1482952019	NM_053056.2:c.108G>T, NP_444284.1:p.Glu36Asp	0	Benign	0.30	Tolerant	-0.49	Neutral	0.897	Disease	0.546	Disease	Benign
21	rs1263446681	NM_053056.2:c.110C>G, NP_444284.1:p.Thr37Ser	0.001	Benign	0.02	Damaging	-0.73	Neutral	0.645	Disease	0.276	Neutral	Benign
22	rs780366497	NM_053056.2:c.114C>G, NP_444284.1:p.Cys38Trp	0.989	Probably Damaging	0.03	Damaging	-0.53	Neutral	0.959	Disease	0.600	Disease	Highly Damaging
23	rs749614691	NM_053056.2:c.122C>T, NP_444284.1:p.Ser41Leu	0.865	Possibly damaging	0.14	Tolerant	-3.23	Deleterious	0.825	Disease	0.574	Disease	Highly Damaging
24	rs1173840555	NM_053056.2:c.135C>A, NP_444284.1:p.Phe45Leu	0.09	Benign	0.06	Tolerant	-4.61	Deleterious	0.690	Disease	0.302	Neutral	Benign
25	rs747665638	NM_053056.2:c.136A>G, NP_444284.1:p.Lys46Glu	0.012	Benign	0.25	Tolerant	-1.27	Neutral	0.373	Neutral	NA	Unclassified	Benign
26	rs1332727287	NM_053056.2:c.148A>G, NP_444284.1:p.Lys50Glu	0	Benign	0.11	Tolerant	-1.38	Neutral	0.826	Disease	0.311	Neutral	Benign
27	rs772785280	NM_053056.2:c.153G>C, NP_444284.1:p.Glu51Asp	0	Benign	0.08	Tolerant	0.33	Neutral	0.410	Neutral	0.219	Neutral	
28	rs760341225	NM_053056.2:c.161C>G, NP_444284.1:p.Pro54Arg	1	Probably Damaging	0.00	Damaging	-6.6	Deleterious	0.928	Disease	0.700	Disease	Highly Damaging
29	rs765904377	NM_053056.2:c.166A>T, NP_444284.1:p.Met56Leu	0.14	Benign	0.09	Tolerant	-2.18	Neutral	0.956	Disease	0.535	Disease	Benign
30	rs748632355	NM_053056.2:c.185C>A, NP_444284.1:p.Thr62Asn,	0.06	Benign	0.45	Tolerant	-1.95	Neutral	0.904	Disease	0.388	Neutral	Benign

		NM_053056.2:c.185C>T, NP_444284.1:p.Thr62Ile	0.854	Possibly damaging	0.32	Tolerant	-2.96	Deleterious	0.955	Disease	0.705	Disease	Highly Damaging
31	rs1299820976	NM_053056.2:c.190A>C, NP_444284.1:p.Met64Leu	0.034	Benign	0.06	Tolerant	-2.23	Neutral	0.731	Disease	0.429	Neutral	Benign
32	rs759345822	NM_053056.2:c.197A>C, NP_444284.1:p.Glu66Ala,	0.981	Probably Damaging	0.00	Damaging	-4.67	Deleterious	0.917	Disease	0.588	Disease	Highly Damaging
		NM_053056.2:c.197A>G, NP_444284.1:p.Glu66Gly	0.994	Probably Damaging	0.00	Damaging	-5.3	Deleterious	0.943	Disease	0.737	Disease	Highly Damaging
33	rs769045064	NM_053056.2:c.206A>G, NP_444284.1:p.Glu69Gly	0.055	Benign	0.03	Damaging	-5.99	Deleterious	0.882	Disease	0.721	Disease	Highly Damaging
34	rs1381472659	NM_053056.2:c.220G>A, NP_444284.1:p.Glu74Lys	0.415	Benign	0.03	Damaging	-3.15	Deleterious	0.853	Disease	0.444	Neutral	Highly Damaging
35	rs1171564865	NM_053056.2:c.223G>A, NP_444284.1:p.Glu75Lys	0.224	Benign	0.01	Damaging	-3.09	Deleterious	0.932	Disease	0.528	Disease	Highly Damaging
36	rs778080996	NM_053056.2:c.226G>A, NP_444284.1:p.Glu76Lys	0.782	Possibly damaging	0.07	Tolerant	-3.25	Deleterious	0.979	Disease	0.495	Neutral	Highly Damaging
37	rs746530862	NM_053056.2:c.228G>C, NP_444284.1:p.Glu76Asp	0.001	Benign	0.55	Tolerant	-1.08	Neutral	0.842	Disease	0.399	Neutral	Benign
38	rs781047821	NM_053056.2:c.235C>T, NP_444284.1:p.Pro79Ser	0.851	Possibly damaging	0.01	Damaging	-6.25	Deleterious	0.864	Disease	0.483	Neutral	Highly Damaging
39	rs1188165799	NM_053056.2:c.241G>A, NP_444284.1:p.Ala81Thr	0.995	Possibly damaging	0.05	Tolerant	-2.79	Deleterious	0.913	Disease	0.885	Disease	Highly Damaging
40	rs1471534715	NM_053056.2:c.244A>G, NP_444284.1:p.Met82Val	0.002	Benign	0.41	Tolerant	-2.75	Deleterious	0.513	Disease	0.321	Neutral	Benign
41	rs1359042394	NM_053056.2:c.277C>A, NP_444284.1:p.Pro93Ser,	0.01	Benign	0.17	Tolerant	-5.11	Deleterious	0.514	Disease	0.233	Neutral	Benign
		NM_053056.2:c.277C>T, NP_444284.1:p.Pro93Thr	0.004	Benign	0.14	Tolerant	-5.37	Deleterious	0.577	Disease	0.489	Neutral	Benign
42	rs1419585038	NM_053056.2:c.278C>T, NP_444284.1:p.Pro93Leu	0.164	Benign	0.01	Damaging	-7.15	Deleterious	0.828	Disease	0.617	Disease	Highly Damaging
43	rs774091629	NM_053056.2:c.280G>T, NP_444284.1:p.Val94Leu	0	Benign	0.39	Tolerant	-0.84	Neutral	0.746	Disease	0.245	Neutral	Benign
44	rs760878237	NM_053056.2:c.285A>C, NM_053056.2:c.285A>G,	0.011	Benign	0.18	Tolerant	-2.52	Deleterious	0.729	Disease	0.342	Neutral	Benign

		NP_444284.1:p.Lys95Asn											
45	rs148113872	NM_053056.2:c.288G>C, NM_053056.2:c.288G>T, NP_444284.1:p.Lys96Asn	0.999	Probably Damaging	0.00	Damaging	-4.23	Deleterious	0.957	Disease	0.673	Disease	Highly Damaging
46	rs140967247	NM_053056.2:c.292C>A, NP_444284.1:p.Arg98Ser	0.195	Benign	0.41	Tolerant	-2.13	Neutral	0.812	Disease	0.341	Neutral	Benign
47	rs1480433568	NM_053056.2:c.302T>G, NP_444284.1:p.Leu101Arg	0.998	Probably Damaging	0.00	Damaging	-5.05	Deleterious	0.989	Disease	0.982	Disease	Highly Damaging
48	rs753115532	NM_053056.2:c.310G>T, NP_444284.1:p.Ala104Ser,	0.987	Probably	0.22	Tolerant	-1.79	Neutral	0.902	Disease	0.498	Neutral	Benign
		NM_053056.2:c.310G>A, NP_444284.1:p.Ala104Thr	0.999	Damaging	0.26	Tolerant	-2.46	Neutral	0.869	Disease	0.564	Disease	Highly Damaging
49	rs764727750	NM_053056.2:c.319A>C, NP_444284.1:p.Met107Leu	0.003	Benign	0.49	Tolerant	-1.84	Neutral	0.391	Neutral	0.417	Neutral	Benign
50	rs751995000	NM_053056.2:c.324C>A, NP_444284.1:p.Phe108Leu	0.104	Benign	0.48	Tolerant	-2.9	Neutral	0.774	Disease	0.266	Neutral	Benign
51	rs374998781	NM_053056.2:c.341A>G, NP_444284.1:p.Lys114Arg	0.048	Benign	0.50	Tolerant	-1.21	Deleterious	0.746	Disease	0.240	Neutral	Benign
52	rs755854763	NM_053056.2:c.350T>A, NP_444284.1:p.Ile117Asn	0.09	Benign	0.57	Tolerant	-2.25	Neutral	0.893	Disease	0.387	Neutral	Benign
53	rs1176401077	NM_053056.2:c.351C>G, NP_444284.1:p.Ile117Met	0.059	Benign	0.14	Tolerant	-0.57	Neutral	0.893 0.883	Disease	0.387 0.622	Neutral Disease	Benign
54	rs1469991905	NM_053056.2:c.361G>A, NP_444284.1:p.Ala121Thr	0.909	Possibly damaging	0.10	Tolerant	-2.26	Neutral	0.915	Disease	0.601	Disease	Highly Damaging
55	rs768334579	NM_053056.2:c.364G>A, NP_444284.1:p.Glu122Lys	0.787	Possibly damaging	0.02	Damaging	-2.95	Neutral	0.930	Disease	0.517	Disease	Highly Damaging
56	rs1373598994	NM_053056.2:c.365A>G, NP_444284.1:p.Glu122Gly	0.289	Benign	0.01	Damaging	-5.42	Deleterious	0.808	Disease	0.530	Disease	Highly Damaging
57	rs1449765247	NM_053056.2:c.367A>C, NP_444284.1:p.Lys123Gln	0.746	Possibly damaging	0.05	Damaging	-3.14	Deleterious	0.608	Disease	0.373	Neutral	Highly Damaging
58	rs774409065	NM_053056.2:c.368A>T, NP_444284.1:p.Lys123Met	0.592	Possibly damaging	0.01	Damaging	-4.79	Deleterious	0.834	Disease	0.819	Disease	Highly Damaging
59	rs748059951	NM_053056.2:c.371T>C, NP_444284.1:p.Leu124Pro	1	Probably Damaging	0.00	Damaging	-6.28	Deleterious	0.989	Disease	0.974	Disease	Highly Damaging
60	rs1329584649	NM_053056.2:c.376A>C, NP_444284.1:p.Ile126Leu	0.003	Benign	0.14	Tolerant	-1.31	Deleterious	0.895	Disease	0.509	Disease	Highly Damaging
61	rs1375845153	NM_053056.2:c.379T>C,	0.999	Probably	0.01	Damaging	-4.35	Neutral	0.946	Disease	0.855	Disease	Highly

		NP_444284.1:p.Tyr127His		Damaging									Damaging
62	rs1312761797	NM_053056.2:c.382A>G, NP_444284.1:p.Thr128Ala	0.892	Possibly damaging	0.01	Damaging	-4.17	Deleterious	0.837	Disease	0.395	Neutral	Highly Damaging
63	rs771952150	NM_053056.2:c.385G>A, NP_444284.1:p.Asp129Asn	0.951	Possibly damaging	0.01	Damaging	-4.29	Deleterious	0.960	Disease	0.670	Disease	Highly Damaging
64	rs1050971	NM_053056.2:c.388A>G, NP_444284.1:p.Asn130Asp,	0.194	Benign	0.06	Tolerant	-3.65	Deleterious	0.923	Disease	0.916	Disease	Highly Damaging
		NM_053056.2:c.388A>T, NP_444284.1:p.Asn130Tyr	0.31	Benign	0.12	Tolerant	-4.91	Deleterious	0.648	Disease	0.257	Neutral	Benign
65	rs1131439	NM_053056.2:c.389A>G, NP_444284.1:p.Asn130Ser	0.196	Benign	0.10	Tolerant	-3.52	Deleterious	0.876	Disease	0.926	Disease	Highly Damaging
66	rs759551813	NM_053056.2:c.392C>T, NP_444284.1:p.Ser131Phe	0.982	Probably Damaging	0.00	Damaging	-4.95	Deleterious	0.956	Disease	0.691	Disease	Highly Damaging
67	rs866931401	NM_053056.2:c.398G>A, NP_444284.1:p.Arg133Gln,	0	Benign	0.30	Tolerant	-0.96	Neutral	0.794	Disease	0.382	Neutral	Benign
		NM_053056.2:c.398G>T, NP_444284.1:p.Arg133Leu	0	Benign	0.15	Tolerant	-2.19	Neutral	0.934	Disease	0.548	Disease	Benign
68	rs765498502	NM_053056.2:c.400C>T, NP_444284.1:p.Pro134Ser	0.185	Benign	0.57	Tolerant	-3.62	Deleterious	0.718	Disease	0.490	Neutral	Benign
69	rs763443319	NM_053056.2:c.403G>A, NP_444284.1:p.Glu135Lys	0.001	Benign	0.50	Tolerant	-1.2	Neutral	0.781	Disease	0.427	Neutral	Benign
70	rs764335132	NM_053056.2:c.404A>T, NP_444284.1:p.Glu135Val	0.003	Benign	0.10	Tolerant	-3.74	Deleterious	0.882	Disease	0.538	Disease	Highly Damaging
71	rs1184664065	NM_053056.2:c.407A>C, NP_444284.1:p.Glu136Ala	0.751	Possibly damaging	0.01	Tolerant	-4.63	Deleterious	0.896	Disease	0.490	Neutral	Benign
72	rs752157015	NM_053056.2:c.409C>A, NP_444284.1:p.Leu137Met	0.99	Probably Damaging	0.03	Damaging	-1.66	Neutral	0.935	Disease	0.823	Disease	Highly Damaging
73	rs1268871232	NM_053056.2:c.417A>C, NP_444284.1:p.Gln139His	0.001	Benign	0.13	Tolerant	-1.96	Neutral	0.806	Disease	0.669	Disease	Benign
74	rs1260695584	NM_053056.2:c.427C>T, NP_444284.1:p.Leu143Phe	0.322	Benign	0.05	Damaging	-3.07	Deleterious	0.832	Disease	0.824	Disease	Highly Damaging
75	rs766855822	NM_053056.2:c.428T>C, NP_444284.1:p.Leu143Pro	0.544	Possibly damaging	0.00	Damaging	-5.82	Deleterious	0.969	Disease	0.937	Disease	Highly Damaging
76	rs753296773	NM_053056.2:c.430C>G,	0.004	Benign	1.00	Tolerant	1.20	Neutral	0.254	Neutral	0.556	Disease	Benign

		NP_444284.1:p.Leu144Val											
77	rs374420164	NM_053056.2:c.442C>T, NP_444284.1:p.Leu148Phe	1	Probably Damaging	0.00	Damaging	-3.84	Deleterious	0.952	Disease	0.943	Disease	Highly Damaging
78	rs1203615889	NM_053056.2:c.454C>A, NP_444284.1:p.Leu152Met	0.992	Probably Damaging	0.08	Tolerant	-1.51	Neutral	0.641	Disease	0.284	Neutral	Benign
79	rs752345237	NM_053056.2:c.460G>C, NP_444284.1:p.Ala154Pro,	0.994	Probably Damaging	0.01	Damaging	-4.02	Deleterious	0.950	Disease	0.490	Neutral	Highly Damaging
		NM_053056.2:c.460G>T, NP_444284.1:p.Ala154Ser	0.087	Benign	0.43	Tolerant	-1.39	Neutral	0.679	Disease	0.235	Neutral	Benign
80	rs1292226646	NM_053056.2:c.463A>G, NP_444284.1:p.Met155Val	0.001	Benign	1.00	Tolerant	1.09	Neutral	0.460	Neutral	0.691	Disease	Benign
81	rs11263523	NM_053056.2:c.474C>A, NM_053056.2:c.474C>G, NP_444284.1:p.His158Gln	0.093	Benign	0.20	Tolerant	-4.34	Deleterious	0.950	Disease	0.685	Disease	Highly Damaging
82	rs781118003	NM_053056.2:c.475G>A, NP_444284.1:p.Asp159Asn	0.998	Probably Damaging	0.01	Damaging	-4.67	Deleterious	0.956	Disease	0.747	Disease	Highly Damaging
83	rs749340325	NM_053056.2:c.476A>G, NP_444284.1:p.Asp159Gly	0.993	Probably Damaging	0.00	Damaging	-6.59	Deleterious	0.915	Disease	0.648	Disease	Highly Damaging
84	rs1447014281	NM_053056.2:c.483T>G, NP_444284.1:p.Ile161Met	0.925	Possibly damaging	0.01	Damaging	-1.76	Neutral	0.937	Disease	0.699	Disease	Highly Damaging
85	rs1332977808	NM_053056.2:c.485A>G, NP_444284.1:p.Glu162Gly	0.304	Benign	0.06	Tolerant	-4.4	Deleterious	0.947	Disease	0.709	Disease	Highly Damaging
86	rs768767169	NM_053056.2:c.488A>G, NP_444284.1:p.His163Arg	0.927	Possibly damaging	0.02	Damaging	-5.56	Deleterious	0.904	Disease	0.811	Disease	Highly Damaging
87	rs537363548	NM_053056.2:c.490T>C, NP_444284.1:p.Phe164Leu	0.206	Benign	0.19	Tolerant	-0.87	Neutral	0.882	Disease	0.273	Neutral	Benign
88	rs1432185835	NM_053056.2:c.493C>G, NP_444284.1:p.Leu165Val	0.723	Possibly damaging	0.05	Tolerant	-2.54	Deleterious	0.882	Disease	0.511	Disease	Highly Damaging
89	rs1261290754	NM_053056.2:c.505C>T, NP_444284.1:p.Pro169Ser	0.167	Benign	0.21	Tolerant	-4.61	Deleterious	0.500	Disease	0.299	Neutral	Benign
90	rs548963461	NM_053056.2:c.509A>T, NP_444284.1:p.Glu170Val	0	Benign	0.58	Tolerant	2.49	Neutral	0.582	Disease	0.237	Neutral	Benign
91	rs747336419	NM_053056.2:c.510G>C, NP_444284.1:p.Glu170Asp	0	Benign	0.07	Tolerant	-0.86	Neutral	0.680	Disease	0.293	Neutral	Benign
92	rs1373641771	NM_053056.2:c.512C>T,	0.003	Benign	0.57	Tolerant	-1.21	Neutral	0.766	Disease	0.335	Neutral	Benign

		NP_444284.1:p.Ala171Val											
93	rs374062310	NM_053056.2:c.522C>G, NP_444284.1:p.Asn174Lys	0.01	Benign	1.00	Tolerant	-0.36	Neutral	0.586	Disease	NA	Unclassified	Benign
94	rs1175448240	NM_053056.2:c.523A>G, NP_444284.1:p.Lys175Glu	0.002	Benign	1.00	Tolerant	-1.43	Neutral	0.503	Disease	NA	Unclassified	Benign
95	rs777185874	NM_053056.2:c.526C>G, NP_444284.1:p.Gln176Glu	0	Benign	1.00	Tolerant	-0.42	Neutral	0.474	Neutral	NA	Unclassified	Benign
96	rs1433903280	NM_053056.2:c.527A>T, NP_444284.1:p.Gln176Leu	0	Benign	0.32	Tolerant	-2.41	Neutral	0.392	Neutral	NA	Unclassified	Benign
97	rs143479406	NM_053056.2:c.536G>A, NP_444284.1:p.Arg179His,	0.919	Possibly damaging	0.05	Tolerant	-3.84	Deleterious	0.907	Disease	NA	Unclassified	Highly Damaging
		NM_053056.2:c.536G>T, NP_444284.1:p.Arg179Leu	0.947	Possibly damaging	0.02	Damaging	-5.10	Deleterious	0.837	Disease	0.268	Neutral	Highly Damaging
98	rs1176469241	NM_053056.2:c.543C>A, NP_444284.1:p.His181Gln	0.903	Possibly damaging	0.03	Damaging	-7.39	Deleterious	0.898	Disease	0.426	Neutral	Highly Damaging
99	rs1434771092	NM_053056.2:c.547C>A, NP_444284.1:p.Gln183Lys	0.237	Benign	0.04	Damaging	-3.08	Deleterious	0.881	Disease	0.508	Disease	Highly Damaging
100	rs1173908293	NM_053056.2:c.551C>G, NP_444284.1:p.Thr184Ser	0.103	Benign	0.08	Tolerant	-3.53	Deleterious	0.766	Disease	0.423	Neutral	Benign
101	rs763697017	NM_053056.2:c.556G>A, NP_444284.1:p.Val186Ile	0.002	Benign	1.00	Tolerant	0.29	Neutral	0.346	Neutral	0.284	Neutral	Benign
102	rs1252289930	NM_053056.2:c.560C>T, NP_444284.1:p.Ala187Val	0.912	Possibly damaging	0.06	Tolerant	-3.58	Deleterious	0.940	Disease	0.440	Neutral	Highly Damaging
103	rs751315372	NM_053056.2:c.562C>T, NP_444284.1:p.Leu188Phe	1	Probably Damaging	0.00	Damaging	-3.76	Deleterious	0.973	Disease	0.838	Disease	Highly Damaging
104	rs757229078	NM_053056.2:c.566G>A, NP_444284.1:p.Cys189Tyr	1	Probably Damaging	0.00	Damaging	-9.72	Deleterious	0.982	Disease	0.706	Disease	Highly Damaging
105	rs534553548	NM_053056.2:c.568G>T, NP_444284.1:p.Ala190Ser	0.894	Possibly damaging	0.12	Tolerant	-2.2	Neutral	0.942	Disease	0.635	Disease	Highly Damaging
106	rs201012923	NM_053056.2:c.577G>A, NP_444284.1:p.Val193Met	0.013	Benign	0.14	Tolerant	-0.55	Neutral	0.871	Disease	0.415	Neutral	Benign
107	rs1282637304	NM_053056.2:c.582G>T, NP_444284.1:p.Lys194Asn	0.02	Benign	0.34	Tolerant	-2.21	Neutral	0.841	Disease	0.485	Neutral	Benign
108	rs759570740	NM_053056.2:c.593A>G, NP_444284.1:p.Asn198Ser	0.02	Benign	0.60	Tolerant	-1.22	Neutral	0.769	Disease	0.386	Neutral	Benign
109	rs1222442441	NM_053056.2:c.596C>T, NP_444284.1:p.Pro199Leu	0.998	Probably Damaging	0.00	Damaging	-7.42	Deleterious	0.776	Disease	0.377	Neutral	Highly Damaging

110	rs1353308163	NM_053056.2:c.598C>T, NP_444284.1:p.Pro200Ser	1	Probably Damaging	0.00	Damaging	-7.2	Deleterious	0.969	Disease	0.831	Disease	Highly Damaging
111	rs893823618	NM_053056.2:c.607G>T, NP_444284.1:p.Val203Leu	0.004	Benign	0.07	Tolerant	-1.29	Neutral	0.853	Disease	0.357	Neutral	Benign
112	rs751664590	NM_053056.2:c.610G>A, NP_444284.1:p.Ala204Thr	1	Probably Damaging	0.00	Damaging	-3.6	Deleterious	0.974	Disease	0.895	Disease	Highly Damaging
113	rs1193747086	NM_053056.2:c.613G>T, NP_444284.1:p.Ala205Ser	0.996	Probably Damaging	0.09	Tolerant	-1.29	Neutral	0.959	Disease	0.756	Disease	Highly Damaging
114	rs1271260640	NM_053056.2:c.614C>T, NP_444284.1:p.Ala205Val	1	Probably Damaging	0.01	Damaging	-2.42	Neutral	0.957	Disease	0.689	Disease	Highly Damaging
115	rs1011426441	NM_053056.2:c.622G>A, NP_444284.1:p.Val208Met	0.977	Probably Damaging	0.09	Tolerant	-1.82	Neutral	0.781	Disease	0.379	Neutral	Benign
116	rs750639632	NM_053056.2:c.626T>C, NP_444284.1:p.Val209Ala	0	Benign	1.00	Tolerant	0.72	Neutral	0.697	Disease	0.325	Neutral	Benign
117	rs756509430	NM_053056.2:c.631G>A, NP_444284.1:p.Ala211Thr	1	Probably Damaging	0.00	Damaging	-3.47	Deleterious	0.968	Disease	0.706	Disease	Highly Damaging
118	rs112525097	NM_053056.2:c.648C>G, NP_444284.1:p.Asn216Lys	0	Benign	0.63	Tolerant	-1.27	Neutral	0.863	Disease	0.376	Neutral	Benign
119	rs149457002	NM_053056.2:c.656G>A, NP_444284.1:p.Ser219Asn	0	Benign	0.52	Tolerant	-1.22	Neutral	0.876	Disease	0.289	Neutral	Benign
120	rs1404273153	NM_053056.2:c.658C>T, NP_444284.1:p.Pro220Ser	0	Benign	0.89	Tolerant	0.44	Neutral	0.805	Disease	NA	Unclassified	Benign
121	rs747703578	NM_053056.2:c.659C>T, NP_444284.1:p.Pro220Leu	0	Benign	0.35	Tolerant	-1.96	Neutral	0.818	Disease	NA	Unclassified	Benign
122	rs1412188214	NM_053056.2:c.661A>C, NP_444284.1:p.Asn221His	0.159	Benign	0.12	Tolerant	-2.37	Neutral	0.920	Disease	0.652	Disease	Benign
123	rs746640562	NM_053056.2:c.670C>A, NP_444284.1:p.Leu224Met	0.034	Benign	0.22	Tolerant	-0.36	Neutral	0.918	Disease	0.649	Disease	Benign
124	rs776600964	NM_053056.2:c.677A>G, NP_444284.1:p.Tyr226Cys,	0	Benign	0.18	Tolerant	-1.42	Neutral	0.853	Disease	0.235	Neutral	Benign
		NM_053056.2:c.677A>T, NP_444284.1:p.Tyr226Phe	0.002	Benign	0.70	Tolerant	-0.93	Neutral	0.816	Disease	0.262	Neutral	Benign
125	rs1222446202	NM_053056.2:c.682C>T, NP_444284.1:p.Arg228Cys	0.004	Benign	0.05	Tolerant	-3.18	Deleterious	0.754	Disease	0.203	Neutral	Benign
126	rs200179137	NM_053056.2:c.683G>A, NP_444284.1:p.Arg228His	0	Benign	0.13	Tolerant	-1.33	Neutral	0.909	Disease	0.572	Disease	Benign

127	rs984643266	NM_053056.2:c.689C>T, NP_444284.1:p.Thr230Ile,	0.476	Possibly damaging	0.60	Tolerant	-2.52	Deleterious	0.940	Disease	0.570	Disease	Highly Damaging
		NM_053056.2:c.689C>A, NP_444284.1:p.Thr230Lys	0.998	Probably Damaging	0.06	Tolerant	-3.71	Deleterious	0.934	Disease	0.255	Neutral	Highly Damaging
128	rs745779714	NM_053056.2:c.691C>T, NP_444284.1:p.Arg231Cys,	0.495	Possibly damaging	0.03	Damaging	-3.35	Deleterious	0.889	Disease	0.410	Neutral	Highly Damaging
		NM_053056.2:c.691C>G, NP_444284.1:p.Arg231Gly,	0	Benign	0.20	Tolerant	-1.83	Neutral	0.860	Disease	0.325	Neutral	Benign
		NM_053056.2:c.691C>A, NP_444284.1:p.Arg231Ser	0	Benign	0.40	Tolerant	-0.58	Neutral	0.838	Disease	0.403	Neutral	Benign
129	rs1228811654	NM_053056.2:c.697C>A, NP_444284.1:p.Leu233Ile	0.998	Probably Damaging	0.05	Tolerant	-1.63	Neutral	0.886	Disease	0.507	Disease	Highly Damaging
130	rs1165177408	NM_053056.2:c.700T>A, NP_444284.1:p.Ser234Thr	0.001	Benign	0.89	Tolerant	-1.08	Neutral	0.716	Disease	0.479	Neutral	Benign
131	rs1423874357	NM_053056.2:c.712A>C, NP_444284.1:p.Lys238Gln	0	Benign	0.23	Tolerant	-1.09	Neutral	0.723	Disease	0.506	Disease	Benign
132	rs760907398	NM_053056.2:c.714G>A, NM_053056.2:c.714G>C, NP_444284.1:p.Lys238Asn	0	Benign	0.89	Tolerant	0.2	Neutral	0.777	Disease	0.547	Disease	Benign
133	rs777225097	NM_053056.2:c.722C>T, NP_444284.1:p.Pro241Leu	0.001	Benign	0.33	Tolerant	-2.98	Deleterious	0.507	Disease	0.265	Neutral	Benign
134	rs913470506	NM_053056.2:c.733C>T, NP_444284.1:p.Arg245Trp	1	Probably Damaging	0.00	Damaging	-5.99	Deleterious	0.933	Disease	0.875	Disease	Highly Damaging
135	rs925960764	NM_053056.2:c.736G>T, NP_444284.1:p.Ala246Ser	0.008	Benign	0.23	Tolerant	-1.31	Neutral	0.797	Disease	0.373	Neutral	Benign
136	rs1417190579	NM_053056.2:c.740G>T, NP_444284.1:p.Cys247Phe,	1	Probably Damaging	0.00	Damaging	-9.74	Deleterious	0.956	Disease	0.310	Neutral	Highly Damaging
		NM_053056.2:c.740G>A, NP_444284.1:p.Cys247Tyr	1	Probably Damaging	0.00	Damaging	-9.63	Deleterious	0.980	Disease	0.427	Neutral	Highly Damaging
137	rs1446903472	NM_053056.2:c.748C>A, NP_444284.1:p.Gln250Lys	0.92	Possibly damaging	0.02	Damaging	-3.28	Deleterious	0.878	Disease	0.246	Neutral	Highly Damaging

138	rs1377678585	NM_053056.2:c.749A>G, NP_444284.1:p.Gln250Arg	0.516	Possibly damaging	0.11	Tolerant	-3.24	Deleterious	0.782	Disease	0.245	Neutral	Highly Damaging
139	rs1173812710	NM_053056.2:c.754G>A, NP_444284.1:p.Glu252Lys	1	Probably Damaging	0.00	Damaging	-3.58	Deleterious	0.946	Disease	0.313	Neutral	Highly Damaging
140	rs750012493	NM_053056.2:c.773G>C, NP_444284.1:p.Ser258Thr	0.23	Benign	0.01	Damaging	-2.06	Deleterious	0.752	Disease	0.180	Neutral	Highly Damaging
141	rs755986542	NM_053056.2:c.778C>T, NP_444284.1:p.Arg260Cys	0.262	Benign	0.11	Tolerant	-3.86	Neutral	0.942	Disease	0.706	Disease	Benign
142	rs779733976	NM_053056.2:c.779G>A, NP_444284.1:p.Arg260His	0.99	Probably Damaging	0.13	Tolerant	-2.41	Deleterious	0.907	Disease	0.457	Neutral	Highly Damaging
143	rs749069429	NM_053056.2:c.782A>G, NP_444284.1:p.Gln261Arg,	0.007	Benign	0.07	Tolerant	-2.06	Neutral	0.765	Disease	0.159	Neutral	Benign
		NM_053056.2:c.782A>C, NP_444284.1:p.Gln261Pro	0.004	Benign	0.08	Tolerant	-3.36	Deleterious	0.950	Disease	0.453	Neutral	Benign
145	rs1306960346	NM_053056.2:c.787C>G, NP_444284.1:p.Gln263Glu	0.008	Benign	0.23	Tolerant	-1.29	Neutral	0.816	Disease	0.212	Neutral	Benign
146	rs768582800	NM_053056.2:c.790C>A, NP_444284.1:p.Gln264Lys	0.035	Benign	0.36	Tolerant	-1.89	Neutral	0.841	Disease	0.166	Neutral	Benign
147	rs1225491115	NM_053056.2:c.797T>C, NP_444284.1:p.Met266Thr	0	Benign	0.60	Tolerant	-0.01	Neutral	0.571	Disease	0.184	Neutral	Benign
148	rs377271027	NM_053056.2:c.800A>G, NP_444284.1:p.Asp267Gly	0	Benign	0.40	Tolerant	-1.25	Neutral	0.636	Disease	0.288	Neutral	Benign
149	rs1323012735	NM_053056.2:c.802C>A, NP_444284.1:p.Pro268Thr	0.067	Benign	0.58	Tolerant	-1.43	Neutral	0.595	Disease	0.219	Neutral	Benign
150	rs1252799046	NM_053056.2:c.808G>A, NP_444284.1:p.Ala270Thr	0.033	Benign	0.61	Tolerant	-0.66	Neutral	0.655	Disease	0.223	Neutral	Benign
151	rs770991095	NM_053056.2:c.809C>T, NP_444284.1:p.Ala270Val	0.033	Benign	0.33	Tolerant	-0.75	Neutral	0.357	Neutral	0.246	Neutral	Benign
152	rs759765773	NM_053056.2:c.811G>A, NP_444284.1:p.Ala271Thr	0	Benign	0.77	Tolerant	0.72	Neutral	0.402	Neutral	0.150	Neutral	Benign
153	rs770021448	NM_053056.2:c.812C>A, NP_444284.1:p.Ala271Asp	0	Benign	0.68	Tolerant	-0.58	Neutral	0.681	Disease	0.125	Neutral	Benign
154	rs775723921	NM_053056.2:c.814G>A, NP_444284.1:p.Glu272Lys	0.007	Benign	0.83	Tolerant	0.24	Neutral	0.590	Disease	0.144	Neutral	Benign
155	rs764757265	NM_053056.2:c.821A>T, NP_444284.1:p.Glu274Val	0.001	Benign	0.28	Tolerant	-0.98	Neutral	0.649	Disease	0.378	Neutral	Benign

156	rs752108957	NM_053056.2:c.823G>C, NP_444284.1:p.Glu275Gln	0.005	Benign	0.55	Tolerant	-0.2	Neutral	0.398	Neutral	0.140	Neutral	Benign
157	rs1436128793	NM_053056.2:c.827A>T, NP_444284.1:p.Glu276Val	0.047	Benign	0.15	Tolerant	-1.14	Neutral	0.711	Disease	0.484	Neutral	Benign
158	rs200912411	NM_053056.2:c.836A>C, NP_444284.1:p.Glu279Ala	0	Benign	0.60	Tolerant	-1.06	Neutral	0.383	Neutral	0.188	Neutral	Benign
159	rs749899296	NM_053056.2:c.845A>G, NP_444284.1:p.Asp282Gly,	0.002	Benign	0.11	Tolerant	-1.29	Neutral	0.691	Disease	0.219	Neutral	Benign
		NM_053056.2:c.845A>T, NP_444284.1:p.Asp282Val	0.425	Benign	0.03	Damaging	-2.72	Deleterious	0.648	Disease	0.209	Neutral	Highly Damaging
160	rs755753468	NM_053056.2:c.847C>G, NP_444284.1:p.Leu283Val	0.037	Benign	0.45	Tolerant	-0.80	Neutral	0.387	Neutral	0.146	Neutral	Benign
161	rs754752470	NM_053056.2:c.853T>A, NP_444284.1:p.Cys285Ser	0.001	Benign	1	Tolerant	-0.72	Neutral	0.496	Neutral	0.272	Neutral	Benign
162	rs1168483993	NM_053056.2:c.854G>T, NP_444284.1:p.Cys285Phe	0.28	Benign	0.56	Tolerant	-3.4	Deleterious	0.923	Disease	0.496	Neutral	Benign
163	rs771951669	NM_053056.2:c.865G>A, NP_444284.1:p.Asp289Asn	1	Probably Damaging	0.01	Damaging	-4.27	Deleterious	0.834	Disease	0.184	Neutral	Highly Damaging
164	rs781165229	NM_053056.2:c.867C>A, NM_053056.2:c.867C>T, NP_444284.1:p.Asp289Glu	1	Probably Damaging	0.03	Damaging	-3.35	Deleterious	0.842	Disease	0.203	Neutral	Highly Damaging
165	rs769921935	NM_053056.2:c.868G>C, NP_444284.1:p.Val290Leu,	0.99	Probably Damaging	0.09	Tolerant	-2.19	Neutral	0.821	Disease	0.317	Neutral	Benign
		NM_053056.2:c.868G>A, NP_444284.1:p.Val290Met	0.998	Probably Damaging	0.00	Damaging	-2.29	Neutral	0.788	Disease	0.475	Neutral	Highly Damaging
166	rs1292246537	NM_053056.2:c.871C>T, NP_444284.1:p.Arg291Trp	0.999	Probably Damaging	0.03	Damaging	-4.86	Deleterious	0.875	Disease	0.799	Disease	Highly Damaging
167	rs775768459	NM_053056.2:c.872G>A, NP_444284.1:p.Arg291Gln,	0.325	Benign	0.31	Tolerant	-1.3	Neutral	0.570	Disease	0.368	Neutral	Benign
		NM_053056.2:c.872G>T, NP_444284.1:p.Arg291Leu	0.212	Benign	0.81	Tolerant	-3.88	Deleterious	0.863	Disease	NA	Unclassified	Benign
168	rs535957987	NM_053056.2:c.876C>G, NP_444284.1:p.Asp292Glu	0.994	Probably Damaging	0.15	Tolerant	-2.56	Deleterious	0.785	Disease	0.257	Neutral	Highly damaging

169	rs1225090625	NM_053056.2:c.877G>A, NP_444284.1:p.Val293Met	0.999	Probably Damaging	0.08	Tolerant	-2.07	Neutral	0.709	Disease	0.342	Neutral	Benign

[a] PolyPhen-2= Polymorphism Phenotyping v2, Scores near to 1 are more confidently predicting the SNP to be damaging. http://genetics.bwh.harvard.edu/pph2/

[b] SIFT= Sorting Intolerant From Tolerant, SNP scores near 0.00 are more predicted as damaging. http://sift.bii.a-star.edu.sg/

[c] PROVEAN= Protein Variation Effect Analyzer, cut off -2.5, scores equal to or above this threshold are predicting the SNP as deleterious http://provean.jcvi.org/index.php

[d] SNP & GO= scores equal or above 0.5 are predicting the SNP as diseased http://snps.biofold.org/snps-and-go/snps-and-go.html

[e] PANTHER= PANTHER scores are given along with SNP & GO scores, scores equal or above 0.5 are predicting the SNP as diseased http://snps.biofold.org/snps-and-go/snps-and-go.html

[f] SNPs predicted as damaging by three or more tools are classified as “highly damaging” and selected for further *in silico* analysis (colored as red).

Supplementary Table 3. SNPeffect web server molecular phenotypic prediction

SNP rs ID and amino acid change	TANGO		WALTZ		LIMBO	
	Scores	Prediction	Scores	Prediction	Scores	Prediction
rs557545630 Arg15Ser	-2.17	does not affect the aggregation tendency of cyclin D1	0.08	does not affect the amyloid propensity of cyclin D1	0.00	does not affect the chaperone binding tendency of cyclin D1
rs534553548 Ala190Ser	-141.12	decreases the aggregation tendency of cyclin D1	8.07	does not affect the amyloid propensity of cyclin D1	0.00	does not affect the chaperone binding tendency of cyclin D1
rs535957987 Asp292Glu	0.00	does not affect the aggregation tendency of cyclin D1	0.00	does not affect the amyloid propensity of cyclin D1	0.00	does not affect the chaperone binding tendency of cyclin D1
rs143479406 Arg179His,	-2.01	does not affect the aggregation tendency of cyclin D1	0.10	does not affect the amyloid propensity of cyclin D1	0.00	does not affect the chaperone binding tendency of cyclin D1
rs143479406 Arg179Leu	-2.19	does not affect the aggregation tendency of cyclin D1	5.85	does not affect the amyloid propensity of cyclin D1	-0.20	does not affect the chaperone binding tendency of cyclin D1