Research Paper

# An Exploratory *in silico* Screening of Potentially Deleterious Single Nucleotide Polymorphisms in the Human *CDH1* Gene

**Kanimozhi D[1], Jissin Mathew[2*], Anly Rose Antu[3], Jasmine Steffi A[4] , Arya K[5]**

[1]Division of Biotechnology, School of Engineering and Technology, Karunya Institute of Technology and Sciences, Coimbatore-641114, Tamil Nadu, India.
[2*]Division of Biotechnology, School of Engineering and Technology, Karunya Institute of Technology and Sciences, Coimbatore-641114, Tamil Nadu, India.
[3]Division of Biotechnology, School of Engineering and Technology, Karunya Institute of Technology and Sciences, Coimbatore-641114, Tamil Nadu, India.
[4]Division of Biotechnology, School of Engineering and Technology, Karunya Institute of Technology and Sciences, Coimbatore-641114, Tamil Nadu, India.
[5]Division of Biotechnology, School of Engineering and Technology, Karunya Institute of Technology and Sciences, Coimbatore-641114, Tamil Nadu, India.

## Corresponding Author

**Jissin Mathew**

Email ID: jissin@karunya.edu

**ABSTRACT**

The *Cadherin-1 (CDH1)* gene encodes E-cadherin, a calcium-dependent adhesion protein essential for maintaining epithelial integrity. Genetic variations within *CDH1* are associated with several epithelial cancers, including hereditary diffuse gastric cancer and lobular breast carcinoma. This study aimed to identify and characterize potentially deleterious single nucleotide polymorphisms (SNPs) in *CDH1* using a computational approach. Nonsynonymous SNPs were retrieved from public databases and analyzed using SIFT, PolyPhen-2, CADD, and Mutation Taster to predict their pathogenic potential. Variants consistently classified as damaging were subjected to further physicochemical and structural evaluation. ProtParam was employed to determine molecular weight, isoelectric point, aliphatic index, instability index, and hydropathicity. Inter Pro Scan analysis identified conserved cadherin domains and functional motifs affected by the variants. Structural validation was performed using Ramachandran plot analysis to assess conformational deviations between wild-type and mutant models. The integrated predictions revealed that specific SNPs could disrupt domain stability and calcium-binding efficiency, leading to loss of adhesion and increased disease susceptibility. Overall, this in-silico study provides insight into the structural and functional consequences of *CDH1* polymorphisms, forming a foundation for future experimental validation and precision medicine applications targeting *CDH1*-associated disorders.

**Keywords:** CDH1 gene, SNP analysis, Functional Characterization, Structural Studies

**How to cite this article:** Kanimozhi D, Mathew J, Antu AR, Jasmine Steffi A, Arya K. An Exploratory in silico Screening of Potentially Deleterious Single Nucleotide Polymorphisms in the Human CDH1 Gene. Int J Drug Deliv Technol. 2026;16(35s): 746-759. DOI: 10.25258/ijddt.16.35s.85

## INDRODUCTION

### Biological Role of CDH1 and E-cadherin

The *Cadherin-1 (CDH1)* gene encodes E-cadherin, a calcium-dependent transmembrane glycoprotein that plays a central role in epithelial cell–cell adhesion. E-cadherin is a core component of adherens junctions, where it maintains epithelial polarity, structural integrity, and coordinated intercellular communication [1]. Through its cytoplasmic interactions with catenins, E-cadherin anchors the actin cytoskeleton to the plasma membrane, thereby preserving tissue architecture and preventing inappropriate cellular dissemination [2].

Structurally, E-cadherin comprises five extracellular cadherin repeats (EC1–EC5) responsible for calcium-dependent homophilic interactions, a single transmembrane domain, and a highly conserved cytoplasmic domain that mediates binding to α-, β-, and p120-catenins. This highly ordered structure is essential for maintaining epithelial homeostasis, and perturbations within any of these domains may compromise adhesive function and cellular cohesion [3].

### CDH1 as a Tumor Suppressor in Epithelial Malignancies

The *CDH1* gene is located on chromosome 16q22.1 and functions as a tumor suppressor gene. Germline and somatic alterations in *CDH1* are strongly associated with multiple epithelial cancers, most notably Hereditary Diffuse Gastric Cancer (HDGC) and invasive lobular breast carcinoma [4]. Loss or functional impairment of E-cadherin disrupts adherens junctions, leading to reduced cell–cell adhesion and increased cellular motility.

A hallmark consequence of CDH1 dysfunction is the induction of epithelial-to-mesenchymal transition (EMT), a biological

process characterized by loss of epithelial markers, acquisition of mesenchymal traits, and enhanced invasive potential. EMT contributes significantly to tumor progression, metastasis, and therapeutic resistance, underscoring the importance of preserving E-cadherin function in epithelial tissues[5].

**Single Nucleotide Polymorphisms and Functional Consequences**

Single Nucleotide Polymorphisms (SNPs) represent the most abundant form of genetic variation in the human genome. While many SNPs are functionally neutral, a subset can profoundly affect gene regulation and protein function. Among these, nonsynonymous SNPs (nsSNPs) result in amino acid substitutions that may alter protein stability, folding, or functional interactions[6].

Such substitutions can disrupt conserved residues, modify physicochemical properties, or induce conformational changes that compromise protein integrity. In the context of tumor suppressor genes such as *CDH1*, deleterious nsSNPs may impair E-cadherin–mediated adhesion, thereby predisposing individuals to cancer development and progression. Consequently, systematic identification and characterization of potentially harmful nsSNPs are critical for understanding disease susceptibility and molecular pathogenesis [7].

**Next-Generation Sequencing and Variant Interpretation**

The advent of Next-Generation Sequencing (NGS) technologies has transformed genomic research by enabling rapid, high-throughput, and cost-effective detection of genetic variants. Applications including Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Whole Transcriptome Sequencing (WTS), and epigenomic analyses have greatly expanded the landscape of variant discovery[8].

However, the interpretation of vast numbers of variants generated through NGS remains a major challenge. Computational approaches play a pivotal role in prioritizing variants with potential functional and clinical relevance[9]. In silico analyses provide an efficient strategy to bridge NGS data with biological interpretation, facilitating the identification of candidate variants for further experimental validation and clinical investigation [10].

**In Silico Prediction of Deleterious CDH1 Variants**

To evaluate the functional significance of nsSNPs in the *CDH1* gene, multiple in silico prediction tools are employed, each based on distinct yet complementary principles. SIFT (Sorting Intolerant From Tolerant) predicts the impact of amino acid substitutions based on evolutionary conservation, with deleterious variants typically occurring at highly conserved positions. PolyPhen-2 (Polymorphism Phenotyping v2) assesses the structural and functional impact of substitutions by integrating sequence conservation, protein structural features, and functional annotations[11].

CADD (Combined Annotation Dependent Depletion) provides a comprehensive deleteriousness score by integrating multiple genomic annotations into a machine-learning framework, enabling prioritization of variants likely to have pathogenic consequences. Mutation Taster further evaluates disease-causing potential by incorporating information on evolutionary conservation, splice-site alterations, protein features, and known disease associations. The combined use of these tools enhances prediction accuracy and reduces false-positive interpretations[12].

**Structural and Functional Characterization Using Bioinformatics Tools**

Beyond pathogenicity prediction, structural and physicochemical analyses are essential for understanding the molecular consequences of SNPs. In this study, ProtParam was utilized to assess key physicochemical properties of the E-cadherin protein, including molecular weight, theoretical isoelectric point, instability index, aliphatic index, and hydropathicity. InterProScan was employed to identify conserved domains, functional motifs, and biologically relevant regions affected by SNPs [13].

Furthermore, Ramachandran plot analysis was performed to evaluate stereochemical quality and conformational deviations between wild-type and mutant protein structures. These integrated analyses provide insight into how specific amino acid substitutions may influence protein folding, stability, and functional integrity[14].

**Significance of the Study**

The present study aims to perform a comprehensive in silico screening and characterization of potentially deleterious SNPs in the human *CDH1* gene. By integrating sequence-based prediction, structural evaluation, and functional annotation, this work seeks to elucidate the molecular mechanisms by which CDH1 variants may compromise E-cadherin function.

The findings of this study may contribute to improved understanding of CDH1-associated cancer susceptibility and support the development of molecular diagnostics, risk assessment strategies, and precision medicine approaches. Additionally, this study highlights the significance of computational genomics as an indispensable component of post-NGS variant interpretation.

METHODOLOGY

**Retrieval of CDH1 Gene and SNP Data**

The *CDH1* (Cadherin-1) gene sequence, both nucleotide and protein, was retrieved from the National Center for Biotechnology Information (NCBI) database (Gene ID: 999). To identify genetic variants, Single Nucleotide Polymorphism (SNP) data were obtained from the dbSNP database and cross-validated using the Ensembl Genome Browser (GRCh38 assembly) [15]. Information such as reference SNP IDs (rsIDs), chromosomal positions, nucleotide variations, and resulting amino acid substitutions were recorded. Only nonsynonymous SNPs (nsSNPs) variants that lead to amino acid changes in the E-cadherin protein were selected for further analysis, as these are more likely to influence protein structure and biological function.

**Functional Impact Prediction**

After SNP identification, multiple computational prediction tools were employed to assess the possible effects of amino acid substitutions on E-cadherin function. The workflow included four key algorithms SIFT, PolyPhen-2, CADD, and Mutation Taster each offering a distinct analytical perspective based on evolutionary conservation, physicochemical changes, and molecular annotations. Variants consistently predicted as deleterious by multiple tools were prioritized for further structural and physicochemical characterization[16].

**SIFT**

The SIFT (Sorting Intolerant From Tolerant) algorithm predicts whether an amino acid substitution affects protein function by analyzing sequence homology and evolutionary conservation across multiple species[17]. It assumes that amino acids conserved through evolution are essential for maintaining structural and functional integrity. The tool requires the FASTA sequence of the *CDH1* protein and variant positions as input, producing a score between 0.0 and 1.0. Variants with tolerance index values less than or equal to 0.05 are predicted to be deleterious, indicating that substitutions at highly conserved sites such as calcium-binding residues or cadherin domain interfaces are likely to disrupt E-cadherin function.

**Polyphen**

The PolyPhen-2 (Polymorphism Phenotyping v2) tool evaluates the probable impact of amino acid substitutions on protein structure and activity using a combination of sequence-based features, structural parameters, and evolutionary data. It predicts how a mutation may alter the hydrophobicity, secondary structure, and accessible surface area of a protein residue [18]. The input includes the protein sequence or UniProt accession number along with the specific amino acid substitution, and the output categorizes mutations as *benign*, *possibly damaging*, or *probably damaging*, with probability scores ranging from 0.0 to 1.0. Mutations with scores above 0.85 were considered "probably damaging," indicating a high likelihood of destabilizing the E- cadherin structure or impairing its calcium-dependent adhesion properties.

**CADD**

The CADD (Combined Annotation Dependent Depletion) algorithm integrates over sixty genomic annotations, including conservation metrics and functional genomic features, into a single quantitative score that reflects the deleteriousness of genetic variants. By combining information from multiple annotation layers, CADD assigns a Phred-scaled score to each variant, where higher values correspond to greater predicted pathogenicity[19]. The input variant data, typically in Variant Call Format (VCF), includes chromosomal position and allelic information. Variants with CADD scores above 20 are considered to fall within the top 1% of the most deleterious variants in the human genome. Such variants were prioritized in this study as potentially pathogenic mutations of *CDH1* warranting further structural and functional evaluation.

**Mutation Taster**

Mutation Taster predicts the disease-causing potential of sequence variants by combining evolutionary conservation data, predicted splice-site changes, regulatory alterations, and effects on protein domains[20]. Using the genomic coordinates, reference sequence, and altered allele as input, the algorithm classifies variants as either *disease-causing* or *polymorphism*, providing an associated confidence value for each prediction. Variants identified as "disease- causing" were interpreted as having the potential to impair *CDH1* gene expression or disrupt E-cadherin protein function, which could consequently contribute to reduced cell adhesion and increased cancer susceptibility.

**Physicochemical Characterization**

To assess the effects of SNPs on E-cadherin's biochemical properties, both wild-type and mutant sequences were analyzed using the ProtParam tool available on the ExPASy server. Parameters such as molecular weight, theoretical isoelectric point (pI), aliphatic index, instability index, and grand average of hydropathicity (GRAVY) were calculated. Variations in these indices provide insights into changes in protein solubility, stability, and hydrophobicity induced by mutations, which may alter proper folding and function[21].

**Functional Domain and Motif Analysis**

InterProScan was employed to identify conserved domains and motifs within the *CDH1* protein sequence. This integrated tool combines multiple databases, including Pfam, SMART, PROSITE, and SUPERFAMILY, to detect structural and functional features [22]. Domain annotation allowed the mapping of mutations to specific regions such as cadherin repeats, calcium-binding motifs, and transmembrane segments. Mutations localized within these critical regions were hypothesized to disrupt E-cadherin's adhesive function and intercellular signaling roles.

**Structural Modeling and Validation**

Homology-based modeling was carried out to generate the three-dimensional structures mutant E-cadherin proteins using template-based prediction tools [23]. The Ramachandran plot was constructed using the PROCHECK server to evaluate the stereochemical quality of the models. The proportion of residues within favored, allowed, and disallowed regions was compared to assess conformational deviations introduced by SNPs[24]. Mutant models showing increased residues in disallowed regions were considered structurally destabilized[25].

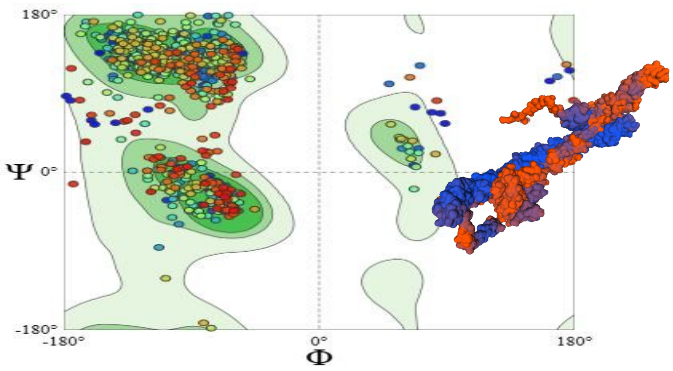



Figure 1: 3D Model of mutated protein and it's structural characteristics analysis with Ramachandra plot

### Data Integration and Interpretation

Finally, results from functional predictions, physicochemical analyses, and structural modeling were integrated to identify the most impactful SNPs within *CDH1*. Variants predicted as damaging by multiple algorithms, positioned within conserved domains, and exhibiting altered physicochemical and structural characteristics were considered potentially pathogenic. These findings form a basis for further experimental validation and potential use of *CDH1* variants as biomarkers for gastric and breast cancer predisposition[26].

## RESULTS

### Variant Effect Predictor

The Variant Effect Predictor (VEP) provided a comprehensive annotation of the SNPs identified within the *CDH1* gene, including their genomic coordinates, nucleotide substitutions, and predicted molecular consequences. Among the total variants screened, a subset was classified as missense mutations, which are most likely to alter protein structure and function [27]. The categorization of SNPs revealed several variants located within coding exons, with potential effects on amino acid sequence and domain integrity. Figure 2 shows the results VEP sheet, which consist of list of SNP found in the *CDH1* gene. These results established the foundation for downstream functional predictions using multiple computational tools.

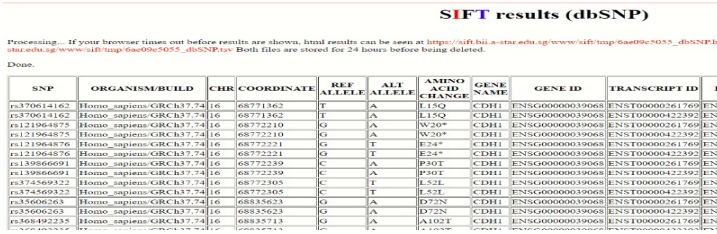
| | #Uploade | Location | Allele | Conseque | IMPACT | SYMBOL | Gene | Feature_t | Feature | BIOTYPE | EXON | INTRON | HGVSc | HGVSp | cDNA_po | CD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | rs155550 | 16:687713 | C | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 125 |
| 3 | rs155550 | 16:687713 | G | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 125 |
| 4 | rs155550 | 16:687713 | T | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 125 |
| 5 | rs155550 | 16:687713 | A | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 126 |
| 6 | rs155550 | 16:687713 | C | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 126 |
| 7 | rs155550 | 16:687713 | G | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 126 |
| 8 | rs878854 | 16:687713 | A | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 127 |
| 9 | rs878854 | 16:687713 | C | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 127 |
| 10 | rs878854 | 16:687713 | T | start_lost | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 127 |
| 11 | rs786201 | 16:687713 | A | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 128 |
| 12 | rs786201 | 16:687713 | C | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 128 |
| 13 | rs786201 | 16:687713 | T | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 128 |
| 14 | rs878854 | 16:687713 | A | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 129 |
| 15 | rs878854 | 16:687713 | T | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 129 |
| 16 | rs106479 | 16:687713 | A | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 131 |
| 17 | rs106479 | 16:687713 | G | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 131 |
| 18 | rs106479 | 16:687713 | T | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 131 |
| 19 | rs587782 | 16:687713 | A | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 132 |
| 20 | rs587782 | 16:687713 | G | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 132 |
| 21 | rs587782 | 16:687713 | T | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 132 |
| 22 | rs196242 | 16:687713 | A | stop_gain | HIGH | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 135 |
| 23 | rs196242 | 16:687713 | C | missense_ | MODERAT | CDH1 | | 999 | Transcript | NM_0043( | protein_cc | | 42370 - | NM_0043( | NP_00435 | 135 |

Figure 2: Variants of CDH1 gene and their annotations

### Functional Impact Prediction

### SIFT

SIFT analysis revealed that several SNPs had scores below the threshold of 0.05, indicating that they are likely to be deleterious. These low-scoring variants suggest substitutions at highly conserved residues, where amino acid changes are not tolerated evolutionarily[26]. Figure 3 illustrates the SIFT prediction on the SNP, which classified the effect of SNP. Such deleterious mutations are expected to disrupt E-cadherin's ability to maintain intercellular adhesion by affecting folding or calcium-binding regions within the cadherin repeats [28]. Therefore, the SIFT analysis highlights critical residues essential for structural stability and functional integrity of E-cadherin, implicating them in disease predisposition[29].




| SNP | ORGANISM/BUILD | CHR | COORDINATE | REF ALLELE | ALT ALLELE | AMINO ACID CHANGE | GENE NAME | GENE ID | TRANSCRIPT ID |
|---|---|---|---|---|---|---|---|---|---|
| rs370614162 | Homo_sapiens/GRCh37.74 | 16 | 68771362 | T | A | L15Q | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs370614162 | Homo_sapiens/GRCh37.74 | 16 | 68771362 | T | A | L15Q | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs121964875 | Homo_sapiens/GRCh37.74 | 16 | 68772210 | G | A | W20* | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs121964875 | Homo_sapiens/GRCh37.74 | 16 | 68772210 | G | A | W20* | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs121964876 | Homo_sapiens/GRCh37.74 | 16 | 68772221 | G | T | E24* | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs121964876 | Homo_sapiens/GRCh37.74 | 16 | 68772221 | G | T | E24* | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs139866691 | Homo_sapiens/GRCh37.74 | 16 | 68772239 | C | A | P30T | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs139866691 | Homo_sapiens/GRCh37.74 | 16 | 68772239 | C | A | P30T | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs374569322 | Homo_sapiens/GRCh37.74 | 16 | 68772305 | C | T | L52L | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs374569322 | Homo_sapiens/GRCh37.74 | 16 | 68772305 | C | T | L52L | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs35606263 | Homo_sapiens/GRCh37.74 | 16 | 68835623 | G | A | D72N | CDH1 | ENSG00000039068 | ENST00000261769 |
| rs35606263 | Homo_sapiens/GRCh37.74 | 16 | 68835623 | G | A | D72N | CDH1 | ENSG00000039068 | ENST00000422392 |
| rs368492235 | Homo_sapiens/GRCh37.74 | 16 | 68835713 | G | A | A102T | CDH1 | ENSG00000039068 | ENST00000261769 |

Figure 3: Functional classification of CDH1 gene using SIFT

### Polyphen

PolyPhen-2 (Polymorphism Phenotyping v2) assesses the potential structural and functional effects of missense mutations using physical and comparative evolutionary parameters. The analyzed *CDH1* variants were classified into three categories: benign, possibly damaging, and probably damaging[30]. A significant number of variants were predicted as "probably damaging", reflecting a high likelihood of altering the protein's stability and interaction capacity. PolyPhen-2 integrates multiple features such as solvent accessibility, hydrophobicity, and secondary structure disruption. Variants predicted as damaging are likely to interfere with E-cadherin's extracellular domain folding, potentially reducing its calcium-dependent adhesive function [26]. These findings complement the SIFT results, reinforcing the evidence that specific substitutions severely compromise protein conformation and stability.

### CADD

CADD (Combined Annotation Dependent Depletion) integrates various genomic features into a single metric to evaluate the deleteriousness of single nucleotide variants[31]. The Phred-scaled CADD scores obtained for *CDH1* SNPs ranged across a spectrum, with several variants scoring above 20, indicating a strong probability of pathogenicity[19]. High CADD scores correlate with functional disruptions that could impact the protein's biochemical behavior and cellular localization. Among these, the variant located at position 2638 (G→A) showed one of the highest Phred scores, signifying its major potential to impair E-cadherin function. These high-impact variants are strongly associated with pathogenic outcomes such as hereditary diffuse gastric cancer (HDGC). Thus, CADD effectively prioritizes SNPs for further validation and provides a quantifiable measure of their biological importance[32].

### Mutation Taster

Mutation Taster integrates evolutionary conservation, splice-site alterations, and protein feature changes to predict whether a variant is disease-causing or benign. The analysis revealed that multiple *CDH1* SNPs were predicted to be "disease-causing", correlating with the predictions from SIFT, PolyPhen, and CADD[33]. These variants likely disrupt essential domains involved in calcium binding and intercellular junction formation. Additionally, Mutation Taster suggested that some SNPs could affect regulatory elements or splicing regions, potentially reducing *CDH1* expression levels. This loss of function aligns with the known molecular mechanism underlying epithelial–mesenchymal transition (EMT) and tumor

progression [34]. The convergence of predictions across all tools strongly supports the hypothesis that deleterious *CDH1* variants contribute to carcinogenic transformation by compromising E-cadherin-mediated adhesion. By analyzing the overall results from various tool, based on the probability can determine the SNP pathogenicity and their significance. Table 1 illustrates the result SNP list which shows deleterious effect and be pathogenic variants of *CDH1* gene which are responsible for potential disease causing[35].

The combined interpretation of these analyses reveals that multiple *CDH1* SNPs, exhibit strong pathogenic potential supported by evolutionary, structural, and functional evidence. Also there are 4 more polymorphism which will cause gastric cancer. These variants likely alter the folding and stability of E-cadherin, weaken cell–cell adhesion, and trigger downstream oncogenic signaling such as Wnt/β-catenin activation. This integrative bioinformatics assessment thus provides valuable insights for early diagnostic marker identification and personalized therapeutic strategies targeting E-cadherin-related cancers. Figure 4 shows the Wild sequence of *CDH1* gene.

**Physicochemical Characterization**

ProtParam analysis of the wild-type and mutant *CDH1* proteins revealed significant alterations in physicochemical properties. The mutated protein exhibited variations in molecular weight, theoretical isoelectric point, and instability index[36]. The increased instability index and altered GRAVY score suggest reduced structural stability and altered hydrophobicity, potentially impairing membrane localization and calcium-dependent adhesion [37]. These physicochemical changes indicate that the mutations may lead to conformational rearrangements that compromise E-cadherin's adhesive efficiency and signaling interactions.

**Functional Domain and Motif Analysis**

This tool have intergrated with Pfam, PRINTS, PROSITE, SMART, PANTHER tools. By comparing the altered sequence against multiple databases, InterProScan functional annotation confirmed that *CDH1* belongs to the classical cadherin family (PR00205) and contains multiple cadherin repeat domains (PF00028) between residues 150–700, a signal peptide (1–20 aa), and a transmembrane helix (709–731 aa). The identified mutations were localized within the cadherin domains responsible for calcium binding, which are critical for maintaining extracellular adhesion. Disruption in these regions may destabilize domain folding, interfere with calcium coordination, and weaken intercellular junctions[38]. Furthermore, these structural disruptions are likely to affect E-cadherin's interaction with β- catenin, leading to aberrant activation of the Wnt/β-catenin signaling pathway. Such dysregulation can promote epithelial–mesenchymal transition (EMT), a hallmark of tumor progression. This mechanistic insight aligns with previous studies linking *CDH1* mutations to hereditary diffuse gastric cancer (HDGC) and invasive lobular breast carcinoma (ILBC) [39].

Integrating the computational findings suggests that specific deleterious SNPs in *CDH1* profoundly impact E-cadherin's structure, physicochemical stability, and biological function. The strong agreement across multiple predictive models supports the reliability of the identified variants as pathogenic [40]. Particularly, the 2638 G→A variant emerges as a highly significant marker for gastric cancer susceptibility. These insights not only contribute to understanding *CDH1*-related disease mechanisms but also open avenues for early genetic screening and personalized therapeutic targeting.

Wild Sequence:

```
>NM_004360.5 Homo sapiens cadherin 1 (CDH1), transcript variant 1, mRNA
AGTGGCGTCGGAACTGCAAAGCACCTGTGAGCTTGCGGAAGTCAGTTCAGACTCCAGCCCGCTCCAGCCCGGCCCGACCCGACCGCACCCGGCGCCTGCCCTCGCTCGGCGTCCCCGGCCAGCCATGGGCCCTTGGAGCC
GCAGCCTCTCGGCGCTGCTGCTGCTGCTGCAGGTCTCCTCTTGGCTCTGCCAGGAGCCGGAGCCCTGCCACCCTGGCTTTGACGCCGAGAGCTACACGTTCACGGTGCCCCGGCGCCACCTGGAGAGAGGCCGCGTCCTG
GGCAGAGTGAATTTTGAAGATTGCACCGGTCGACAAAGGACAGCCTATTTTTCCCTCGACACCCGATTCAAAGTGGGCACAGATGGTGTGATTACAGTCAAAAGGCCTCTACGGTTTCATAACCCACAGATCCATTTCTT
GGTCTACGCCTGGGACTCCACCTACAGAAAGTTTTCCACCAAAGTCACGCTGAATACAGTGGGGCACCACCACCGCCCCCCGCCCCATCAGGCCTCCGTTTCTGGAATCCAAGCAGAATTGCTCACATTTCCCAACTCCT
CTCCTGGCCTCAGAAGACAGAAGAGAGACTGGGTTATTCCTCCCATCAGCTGCCCAGAAAATGAAAAAGGCCCATTTCCTAAAAACCTGGTTCAGATCAAATCCAACAAAGACAAAGAAGGCAAGGTTTTCTACAGCATC
ACTGGCCAAGGAGCTGACACACCCCCTGTTGGTGTCTTTATTATTGAAAGAGAAACAGGATGGCTGAAGGTGACAGAGCCTCTGGATAGAGAACGCATTGCCACATACACTCTCTTCTCTCACGCTGTGTCATCCAACGG
GAATGCAGTTGAGGATCCAATGGAGATTTTGATCACGGTAACCGATCAGAATGACAACAAGCCCGAATTCACCCAGGAGGTCTTTAAGGGGTCTGTCATGGAAGGTGCTCTTCCAGGAACCTCTGTGATGGAGGTCACAG
CCACAGACGCGGACGATGATGTGAACACCTACAATGCCGCCATCGCTTACACCATCCTCAGCCAAGATCCTGAGCTCCCTGACAAAAATATGTTCACCATTAACAGGAACACAGGAGTCATCAGTGTGGTCACCACTGGG
CTGGACCGAGAGAGTTTCCCTACGTATACCCTGGTGGTTCAAGCTGCTGACCTTCAAGGTGAGGGGTTAAGCACAACAGCAACAGCTGTGATCACAGTCACTGACACCAACGATAATCCTCCGATCTTCAATCCCACCAC
GTACAAGGGTCAGGTGCCTGAGAACGAGGCTAACGTCGTAATCACCACACTGAAAGTGACTGATGCTGATGCCCCAATACCCCAGCGTGGGAGGCTGTATACACCATATTGAATGATGATGGTGGACAATTTGTCGTCA
CCACAAATCCAGTGAACAACGATGGCATTTTGAAAACAGCAAAGGGCTTGGATTTTGAGGCCAAGCAGCAGTACATTCTACACGTAGCAGTGACGAATGTGGTACCTTTTGAGGTCTCTCTCACCACCTCCACAGCCACC
GTCACCGTGGATGTGCTGGATGTGAATGAAGCCCCCATCTTTGTGCCTCCTGAAAAGAGAGTGGAAGTGTCCGAGGACTTTGGCGTGGGCCAGGAAATCACATCCTACACTGCCCAGGAGCCAGACACATTTATGGAACA
GAAAATAACATATCGGATTTGGAGAGACACTGCCAACTGGCTGGAGATTAATCCGGACACTGGTGCCATTTCCACTCGGGCTGAGCTGGACAGGGAGGATTTTGAGCACGTGAAGAACAGCACGTACACAGCCCTAATCA
TAGCTACAGACAATGGTTCTCCAGTTGCTACTGGAACAGGGACACTTCTGCTGATCCTGTCTGATGTGAATGACAACGCCCCATACCAGAACCTCGAACTATATTCTTCTGTGAGAGGAATCCAAAGCCTCAGGTCATA
AACATCATTGATGCAGACCTTCCTCCCAATACATCTCCCTTCACAGCAGAACTAACACACGGGGCGAGTGCCAACTGGACCATTCAGTACAACGACCCAACCCAAGAATCTATCATTTTGAAGCCAAAGATGGCCTTAGA
GGTGGGTGACTACAAAATCAATCTCAAGCTCATGGATAACCAGAATAAAGACCAAGTGACCACCTTAGAGGTCAGCGTGTGTGACTGTGAAGGGGCCGCTGGCGTCTGTAGGAAGGCACAGCCTGTCGAAGCAGGATTGC
AAATTCCTGCCATTCTGGGGATTCTTGGAGGAATTCTTGCTTTGCTAATTCTGATTCTGCTGCTCTTGCTGTTTCTTCGGAGGAGAGCGGTGGTCAAAGAGCCCTTACTGCCCCCAGAGGATGACACCCGGGACAACGTT
TATTACTATGATGAAGAAGGAGGCGGAGAAGAGGACCAGGACTTTGACTTGAGCCAGCTGCACAGGGGCCTGGACGCTCGGCCTGAAGTGACTCGTAACGACGTTGCACCAACCCTCATGAGTGTCCCCCGGTATCTTCC
CCGCCCTGCCAATCCCGATGAAATTGGAAATTTTATTGATGAAAATCTGAAAGCGGCTGATACTGACCCCACAGCCCCGCCTTATGATTCTCTGCTCGTGTTTGACTATGAAGGAAGCGGTTCCGAAGCTGCTAGTCTGA
GCTCCCTGAACTCCTCAGAGTCAGACAAAGACCAGGACTATGACTACTTGAACGAATGGGGCAATCGCTTCAAGAAGCTGGCTGACATGTACGGAGGCGGCGAGGACGACTAGGGGACTCGAGAGAGGCGGGCCCCAGAC
CCATGTGCTGGGAAATGCAGAAATCACGTTGCTGGTGGTTTTTCAGCTCCCTTCCCTTGAGATGAGTTTCTGGGGAAAAAAAAGAGACTGGTTAGTGATGCAGTTAGTATAGCTTTATACTCTCTCCACTTTATAGCTCT
AATAAGTTTGTGTTAGAAAAGTTTCGACTTATTTCTTAAAGCTTTTTTTTTTTTCCCATCACTCTTTACATGGTGGTGATGTCCAAAAGATACCCAAATTTTAATATTCCAGAAGAACAACTTTAGCATCAGAAGGTTCA
CCCAGCACCTTGCAGATTTTCTTAAGGAATTTTGTCTCACTTTTAAAAAGAAGGGGAGAAGTCAGCTACTCTAGTTCTGTTGTTTTGTGTATATAATTTTTTAAAAAAAATTTGTGTGCTTCTGCTCATTACTACACTGG
TGTGTCCCTCTGCCTTTTTTTTTTTTTAAGACAGGGTCTCATTCTATCGGCCAGGCTGGAGTGCAGTGGTGCAATCACAGCTCACTGCAGCCTTGTCCTCCCAGGCTCAAGCTATCCTTGCACCTCAGCCTCCCAAGTA
GCTGGGACCACAGGCATGCACCACTACGCATGACTAATTTTTTAAATATTTGAGACGGGGTCTCCCTGTGTTACCCAGGCTGGTCTCAAACTCCTGGGCTCAAGTGATCCTCCCATCTTGGCCTCCCAGAGTATTGGGAT
TACAGACATGAGCCACTGCACCTGCCCAGCTCCCCAACTCCCTGCCATTTTTTAAGAGACAGTTTCGCTCCATCGCCCAGGCCTGGGATGCAGTGATGTGATCATAGCTCACTGTAACCTCAAACTCTGGGGCTCAAGCA
GTTCTCCCACCAGCCTCCTTTTTATTTTTTTGTACAGATGGGGTCTTGCTATGTTGCCCAAGCTGGTCTTAAACTCCTGGCCTCAAGCAATCCTTCTGCCTTGGCCCCCCAAAGTGCTGGGATTGTGGGCATGAGCTGCT
GTGCCCAGCCTCCATGTTTTAATATCAACTCTCACTCCTGAATTCAGTTGCTTTGCCCAAGATAGGAGTTCTCTGATGCAGAAATTATTGGGCTCTTTTAGGGTAAGAAGTTTGTGTCTTTGTCTGGCCACATCTTGACT
AGGTATTGTCTACTCTGAAGACCTTTAATGGCTTCCCTCTTTCATCTCCTGAGTATGTAACTTGCAATGGGCAGCTATCCAGTGACTTGTTCTGAGTAAGTGTGTTCATTAATGTTTATTTAGCTCTGAAGCAAGAGTGA
TATACTCCAGGACTTAGAATAGTGCCTAAAGTGCTGCAGCCAAAGACAGAGCGGAACTATGAAAAGTGGGCTTGGAGATGGCAGGAGAGCTTGTCATTGAGCCTGGCAATTTAGCAAACTGATGCTGAGGATGATTGAGG
TGGGTCTACCTCATCTCTGAAAATTCTGGAAGGAATGGAGGAGTCTCAACATGTGTTTCTGACACAAGATCCGTGGTTTGTACTCAAAGCCCAGAATCCCCAAGTGCCTGCTTTTGATGATGTCTACAGAAAATGCTGGC
TGAGCTGAACACATTTGCCCAATTCCAGGTGTGCACAGAAAACCGAGAATATTCAAAATTCCAAATTTTTTTCTTAGGAGCAAGAAGAAAATGTGGCCCTAAAGGGGGTTAGTTGAGGGGTAGGGGGTAGTGAGGATCTT
GATTTGGATCTCTTTTATTTAAATGTGAATTTCAACTTTTGACAATCAAAGAAAAGACTTTTGTTGAAATAGCTTTACTGTTTCTCAAGTGTTTTGGAGAAAAAAATCAACCCTGCAATCACTTTTTGGAATTGTCTTG
ATTTTTCGGCAGTTCAAGCTATATCGAATATAGTTCTGTGTAGAGAATGTCACTGTAGTTTTGAGTGTATACATGTGTGGGTGCTGATAATTGTGTATTTTCTTTGGGGGTGGAAAAGGAAAACAATTCAAGCTGAGAAA
AGTATTCTCAAAGATGCATTTTTATAAATTTTATTAAACAATTTTGTTAAA
```

Figure 4: Wild type Nucleotide sequence of CDH1 gene

Table 1: Pathogenic Variants of CDH1 gene

| RS ID | NM ID | NP ID | SIFT | POLYPHEN | CADD | Mutation Tester |
|---|---|---|---|---|---|---|
| rs35606263 | NM_004360.5:C.214G>A | NP_004351.1:P.ASP72ASN | Tolerated | Probably Damaging | Deleterious | Benign |
| rs35606263 | NM_004360.5:C.214G>C | NP_004351.1:P.ASP72HIS | Tolerated | Probably Damaging | Deleterious | Benign |
| rs35606263 | NM_004360.5:C.214G>T | NP_004351.1:P.ASP72TYR | Tolerated | Probably Damaging | Deleterious | Benign |
| rs34466743 | NM_004360.5:C.1178T>A | NP_004351.1:P.ILE393ASN | Deleterious | Probably Damaging | Deleterious | Benign |
| rs36087757 | NM_004360.5:C.1417G>A | NP_004351.1:P.VAL473ILE | Tolerated | Possibly Damaging | Deleterious | Benign |
| rs36087757 | NM_004360.5:C.1417G>C | NP_004351.1:P.VAL473LEU | Tolerated | Possibly Damaging | Deleterious | Benign |
| rs35520415 | NM_004360.5:C.1433T>C | NP_004351.1:P.LEU478PRO | Tolerated | Possibly Damaging | Deleterious | Benign |
| rs35187787 | NM_004360.5:C.1774G>A | NP_004351.1:P.ALA592THR | Deleterious | Benign | Deleterious | Benign |
| rs35187787 | NM_004360.5:C.1774G>T | NP_004351.1:P.ALA592SER | Deleterious | Benign | Deleterious | Benign |
| rs33935154 | NM_004360.5:C.1849G>A | NP_004351.1:P.ALA617THR | Tolerated | Benign | Deleterious | Benign |
| rs33935154 | NM_004360.5:C.1849G>C | NP_004351.1:P.ALA617PRO | Tolerated | Benign | Deleterious | Benign |
| rs33935154 | NM_004360.5:C.1849G>T | NP_004351.1:P.ALA617SER | Tolerated | Benign | Deleterious | Benign |
| rs2276331 | NM_004360.5:C.1888C>G | NP_004351.1:P.LEU630VAL | Deleterious | Probably Damaging | Deleterious | Benign |
| rs2276331 | NM_004360.5:C.1888C>T | NP_004351.1:P.LEU630%3D | Deleterious | Probably Damaging | Benign | Benign |
| rs2276331 | NM_004360.5:C.1888C>G | NP_004351.1:P.LEU630VAL | Deleterious | Probably Damaging | Deleterious | Benign |
| rs2276331 | NM_004360.5:C.1888C>T | NP_004351.1:P.LEU630%3D | Deleterious | Probably Damaging | Benign | Benign |
| rs34507583 | NM_004360.5:C.2638G>A | NP_004351.1:P.GLU880LYS | Deleterious | Possibly Damaging | Deleterious | Deleterious |
| rs34507583 | NM_004360.5:C.2638G>C | NP_004351.1:P.GLU880GLN | Deleterious | Possibly Damaging | Deleterious | Benign |

**>NP_004351.1 Wild type cadherin-1 [Homo sapiens]**

MGPWSRSLSALLLLLQVSSWLCQEPEPCHPGFDAESYTF
TVPRRHLERGRVLGRVNFEDCTGRQRTAYFSLDTRFKVG

TDGVITVKRPLRFHNPQIHFLVYAWDSTYRKFSTKVTLNT
VGHHHRPPPHQASVSGIQAELLTFPNSSPGLRRQKRDWVI
PPISCPENEKGPFPKNLVQIKSNKDKEGKVFYSITGQGAD
TPPVGVFIIERETGWLKVTEPLDRERIATYTLFSHAVSSNG
NAVEDPMEILITVTDQNDNKPEFTQEVFKGSVMEGALPG
TSVMEVTATDADDDVNTYNAAIAYTILSQDPELPDKNMF
TINRNTGVISVVTTGLDRESFPTYTLVVQAADLQGEGLST
TATAVITVTDTNDNPPIFNPTTYKGQVPENEANVVITTLKV
TDADAPNTPAWEAVYTILNDDGGQFVVTTNPVNNDGILK
TAKGLDFEAKQQYILHVAVTNVVPFEVSLTTSTATVTVDV
LDVNEAPIFVPPEKRVEVSEDFGVGQEITSYTAQEPDTFM
EQKITYRIWRDTANWLEINPDTGAISTRAELDREDFEHVK
NSTYTALIIATDNGSPVATGTGTLLLILSDVNDNAPIPEPRT
IFFCERNPKPQVINIIDADLPPNTSPFTAELTHGASANWTIQ
YNDPTQESIILKPKMALEVGDYKINLKLMDNQNKDQVTT
LEVSVCDCEGAAGVCRKAQPVEAGLQIPAILGILGGILAL
LILILLLLLFLRRRAVVKEPLLPPEDDTRDNVYYYDEEGG
GEEDQDFDLSQLHRGLDARPEVTRNDVAPTLMSVPRYLP
RPANPDEIGNFIDENLKAADTDPTAPPYDSLLVFDYEGSG
SEAASLSSLNSSESDKDQDYDYLNEWGNRFKKLADMYG
GGEDD

**>NP_004351.1 Mutated cadherin-1 [Homo sapiens]**

MGPWSRSLSALLLLLQVSSWLCQEPEPCHPGFDAESYTF
TVPRRHLERGRVLGRVNFEDCTGRQRTAYFSLDTRFKVG
TDGVITVKRPLRFHNPQIHFLVYAWDSTYRKFSTKVTLNT
VGHHHRPPPHQASVSGIQAELLTFPNSSPGLRRQKRDWVI
PPISCPENEKGPFPKNLVQIKSNKDKEGKVFYSITGQGAD
TPPVGVFIIERETGWLKVTEPLDRERIATYTLFSHAVSSNG
NAVEDPMEILITVTDQNDNKPEFTQEVFKGSVMEGALPG
TSVMEVTATDADDDVNTYNAAIAYTILSQDPELPDKNMF
TINRNTGVISVVTTGLDRESFPTYTLVVQAADLQGEGLST
TATAVITVTDTNDNPPIFNPTTYKGQVPENEANVVNTTLK
VTDADAPNTPAWEAVYTILNDDGGQFVVTTNPVNNDGIL
KTAKGLDFEAKQQYILHVAVTNVVPFEVSLTTSTATVTVD
VLDVNEAPIFVPPEKRVEVSEDFGVGQEITSYTAQEPDTF
MEQKITYRIWRDTANWLEINPDTGAISTRAELDREDFEH
VKNSTYTALIIATDNGSPVATGTGTLLLILSDVNDNAPIPEP
RTIFFCERNPKPQVINIIDADLPPNTSPFTAEVTHGASANW
TIQYNDPTQESIILKPKMALEVGDYKINLKLMDNQNKDQ
VTTLEVSVCDCEGAAGVCRKAQPVEAGLQIPAILGILGGI
LALLILILLLLLFLRRRAVVKEPLLPPEDDTRDNVYYYDE
EGGGEEDQDFDLSQLHRGLDARPEVTRNDVAPTLMSVP
RYLPRPANPDEIGNFIDENLKAADTDPTAPPYDSLLVFDY
EGSGSEAASLSSLNSSESDKDQDYDY
LNEWGNRFKKLADMYGGGKDD

**Figure 5: Amino acid sequence of Cadherin protein between wild type and mutated type**

```
Number of amino acids: 882
Theoretical pI: 4.60
Molecular weight: 97442.13

Amino acid composition:
Ala (A)  59       6.7%
Arg (R)  39       4.4%
Asn (N)  51       5.8%
Asp (D)  68       7.7%
Cys (C)   8       0.9%
Gln (Q)  31       3.5%
Glu (E)  59       6.7%
Gly (G)  54       6.1%
His (H)  13       1.5%
Ile (I)  47       5.3%
Leu (L)  75       8.5%
Lys (K)  36       4.1%
Met (M)  10       1.1%
Phe (F)  33       3.7%
Pro (P)  64       7.3%
Ser (S)  48       5.4%
Thr (T)  80       9.1%
Trp (W)  10       1.1%
Tyr (Y)  26       2.9%
Val (V)  71       8.0%
Pyl (O)   0       0.0%
Sec (U)   0       0.0%

 (B)   0          0.0%
 (Z)   0          0.0%
 (X)   0          0.0%

Total number of negatively charged residues (Asp + Glu): 127
Total number of positively charged residues (Arg + Lys): 75
```

**Figure 6: Physiochemical Properties of Mutated Protein using Prot Param tool**

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 4325 |
| Hydrogen | H | 6747 |
| Nitrogen | N | 1153 |
| Oxygen | O | 1373 |
| Sulfur | S | 18 |

**Formula:** $C_{4325}H_{6747}N_{1153}O_{1373}S_{18}$
**Total number of atoms:** 13616

**Extinction coefficients:**
Extinction coefficients are in units of $M^{-1}$ $cm^{-1}$, at 280 nm measured in water.
Ext. coefficient 94240
Abs 0.1% (=1 g/l) 0.967, assuming all pairs of Cys residues form cystines

Ext. coefficient 93740
Abs 0.1% (=1 g/l) 0.962, assuming all Cys residues are reduced

**Estimated half-life:**
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

**Instability index:**
The instability index (II) is computed to be 35.01
This classifies the protein as stable.

**Aliphatic index:** 83.98
**Grand average of hydropathicity (GRAVY):**-0.360

## DISCUSSION

The present computational analysis systematically evaluated nonsynonymous SNPs within the CDH1 gene to understand their potential structural and functional implications on E-cadherin integrity[34]. Variant annotation revealed several missense substitutions located predominantly within coding exons, suggesting a high probability of direct effects on protein function. Functional prediction tools consistently indicated that multiple variants are deleterious, particularly those affecting conserved residues within extracellular cadherin repeat domains that mediate calcium-dependent cell adhesion[26]. Such substitutions are likely to disrupt calcium coordination, impair homophilic binding between epithelial cells, and ultimately weaken adherens junction stability. The concordance observed among SIFT, PolyPhen-2, CADD, and Mutation Taster strengthens confidence in the predicted pathogenicity and reduces the likelihood of false-positive interpretations, highlighting the reliability of integrative in-silico approaches.

Physicochemical characterization further demonstrated alterations in molecular weight, hydrophobicity, instability index, and theoretical isoelectric point between wild-type and mutant proteins[39]. Increased instability indices and shifts in hydropathicity suggest reduced structural stability and possible misfolding, which may compromise membrane localization and protein–protein interactions. Domain analysis confirmed that several mutations fall within calcium-binding cadherin repeats and near transmembrane regions critical for maintaining epithelial cohesion. Structural deviations observed in modelled proteins, supported by Ramachandran plot validation, indicate conformational strain that could impair adhesive functionality and downstream signalling interactions with catenins. Disruption of E-cadherin–catenin complexes is a known trigger for epithelial–mesenchymal transition, a process strongly associated with tumor invasion and metastasis.

Among the analysed variants, the G→A substitution at position 2638 emerged as particularly significant due to its consistently high deleteriousness scores across prediction platforms, suggesting strong clinical relevance[9]. Variants identified in this study may therefore serve as candidate biomarkers for susceptibility to hereditary diffuse gastric cancer and invasive lobular breast carcinoma, conditions frequently associated with CDH1 dysfunction. Moreover, the integration of functional prediction, structural modelling, and physicochemical assessment provides a comprehensive understanding of how single nucleotide alterations may translate into phenotypic consequences at the molecular level. While computational predictions provide valuable preliminary insights, experimental validation through biochemical assays, expression studies, and clinical correlation remains necessary to confirm pathogenicity[2]. Overall, this study underscores the importance of integrative bioinformatics analyses in prioritizing disease-associated variants and advancing precision medicine strategies targeting CDH1-related malignancies.

## CONCLUSION

This study comprehensively analyzed single nucleotide polymorphisms (SNPs) within the CDH1 gene using multiple computational prediction tools, including SIFT, PolyPhen-2, CADD, and Mutation Taster, complemented by structural and functional characterization through Inter ProScan, Prot Param, and Ramachandran plot analysis. The integrated approach identified several potentially deleterious variants, with the G→A substitution at chromosome 16:2638 emerging as the most pathogenic. Functional prediction tools consistently indicated that this mutation could destabilize E-cadherin's structure, disrupt calcium-dependent adhesion, and impair epithelial integrity, thereby enhancing the risk of tumorigenesis, particularly hereditary diffuse gastric cancer (HDGC) [41]. The structural analyses further revealed alterations in physicochemical parameters such as molecular weight, isoelectric point, instability index, and GRAVY score, indicating decreased protein stability and altered hydrophobic interactions [42]. Domain analysis confirmed that these mutations lie within crucial cadherin repeat regions responsible for calcium binding, suggesting functional interference with β-catenin interaction and downstream Wnt signaling activation [43]. Collectively, these findings underscore the role of CDH1 variants in cancer pathogenesis and emphasize their potential utility as molecular biomarkers for early diagnosis and precision-based therapeutic interventions. The current study is based solely on in-silico analyses; experimental validation through molecular docking, expression studies, or functional assays is required to confirm the predicted pathogenic effects and clinical relevance.

## Reference


1. Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392:402–5. https://doi.org/10.1038/32918

2. Oliveira C, Bordin MC, Grehan N, Huntsman D, Suriano G, Machado JC, et al. Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred. Hum Mutat. 2002;19:510–7. https://doi.org/10.1002/humu.10068

3. Error in Byline. Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. JAMA Oncol. 2015;1:110. https://doi.org/10.1001/jamaoncol.2015.0410

4. Corso G, Figueiredo J, De Angelis S Pietro, Corso F, Girardi A, Pereira J, et al. E-cadherin deregulation in breast cancer. J Cell Mol Med. 2020;24:5930–6. https://doi.org/10.1111/jcmm.15140

5. Coopman P, Djiane A. Adherens Junction and E-Cadherin complex regulation by epithelial polarity. Cellular and Molecular Life Sciences. 2016;73:3535–53. https://doi.org/10.1007/s00018-016-2260-8

6. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi.org/10.1038/nature15393

7. Luo W, Fedda F, Lynch P, Tan D. CDH1 Gene and Hereditary Diffuse Gastric Cancer Syndrome: Molecular and Histological Alterations and Implications for Diagnosis And Treatment. Front Pharmacol. 2018;9. https://doi.org/10.3389/fphar.2018.01421

8. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet. 2017;18:599–612. https://doi.org/10.1038/nrg.2017.52

9. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. 2017;18:225. https://doi.org/10.1186/s13059-017-1353-5

10. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015;17:405–24. https://doi.org/10.1038/gim.2015.30

11. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9. https://doi.org/10.1038/nmeth0410-248

12. Garcia FA de O, Andrade ES de, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. Front Genet. 2022;13. https://doi.org/10.3389/fgene.2022.1010327

13. Parisini E, Higgins JMG, Liu J, Brenner MB, Wang J. The Crystal Structure of Human E-cadherin Domains 1 and 2, and Comparison with other Cadherins in the Context of Adhesion Mechanism. J Mol Biol. 2007;373:401–11. https://doi.org/10.1016/j.jmb.2007.08.011

14. Savojardo C, Babbi G, Martelli PL, Casadio R. Mapping OMIM Disease–Related Variations on Protein Domains Reveals an Association

Among Variation Type, Pfam Models, and Disease Classes. Front Mol Biosci. 2021;8. https://doi.org/10.3389/fmolb.2021.617016

15. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44:D862-8. https://doi.org/10.1093/nar/gkv1222

16. Roy S, Maheshwari N, Chauhan R, Sen NK, Sharma A. Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation. 2011;6:315–9. https://doi.org/10.6026/97320630006315

17. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:W452–7. https://doi.org/10.1093/nar/gks539

18. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc Hum Genet. 2013;76. https://doi.org/10.1002/0471142905.hg0720s76

19. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–94. https://doi.org/10.1093/nar/gky1016

20. Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz JM, Seelow D. MutationTaster2021. Nucleic Acids Res. 2021;49:W446–51. https://doi.org/10.1093/nar/gkab266

21. Sigrist CJA, Cuche BA, de Castro E, Coudert E, Redaschi N, Bridge A. The PROSITE database for protein families, domains, and sites. Nucleic Acids Res. 2025; https://doi.org/10.1093/nar/gkaf1188

22. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33:W116–20. https://doi.org/10.1093/nar/gki442

23. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7:42717. https://doi.org/10.1038/srep42717

24. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46:W296–303. https://doi.org/10.1093/nar/gky427

25. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. Oxford University Press; 2018;46:W296–303. https://doi.org/10.1093/nar/gky427

26. Flanagan SE, Patch A-M, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers. 2010;14:533–7. https://doi.org/10.1089/gtmb.2010.0036

27. Awotoye W, Mossey PA, Hetmanski JB, Gowans LJJ, Eshete MA, Adeyemo WL, et al. Damaging Mutations in *AFDN* Contribute to Risk of Nonsyndromic Cleft Lip With or Without Cleft Palate. The Cleft Palate Craniofacial Journal. 2024;61:697–705. https://doi.org/10.1177/10556656221135926

28. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073–81. https://doi.org/10.1038/nprot.2009.86

29. Ng PC. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–4. https://doi.org/10.1093/nar/gkg509

30. Ahmed I, John P, Bhatti A. Association analysis of Vascular Endothelial Growth Factor-A (VEGF-A) polymorphism in rheumatoid arthritis using computational approaches. Sci Rep. 2023;13:21957. https://doi.org/10.1038/s41598-023-47780-8

31. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–94. https://doi.org/10.1093/nar/gky1016

32. Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res. 2024;52:D1143–54. https://doi.org/10.1093/nar/gkad989

33. Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz JM, Seelow D. MutationTaster2021. Nucleic Acids Res. 2021;49:W446–51. https://doi.org/10.1093/nar/gkab266

34. van der Post RS, Vogelaar IP, Manders P, van der Kolk LE, Cats A, van Hest LP, et al. Accuracy of Hereditary Diffuse Gastric Cancer Testing Criteria and Outcomes in Patients With a Germline Mutation in CDH1. Gastroenterology.

2015;149:897-906.e19. https://doi.org/10.1053/j.gastro.2015.06.003

35. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11:361–2. https://doi.org/10.1038/nmeth.2890

36. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031

37. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571–607. https://doi.org/10.1385/1-59259-890-0:571

38. Wu P-Y. Germline promoter hypermethylation of tumor suppressor genes in gastric cancer. World J Gastroenterol. 2012;18:70. https://doi.org/10.3748/wjg.v18.i1.70

39. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031

40. Gasteiger E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784–8. https://doi.org/10.1093/nar/gkg563

41. Yue P, Li Z, Moult J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. J Mol Biol. 2005;353:459–73. https://doi.org/10.1016/j.jmb.2005.08.020

42. Vendome J, Felsovalyi K, Song H, Yang Z, Jin X, Brasch J, et al. Structural and energetic determinants of adhesive binding specificity in type I cadherins. Proceedings of the National Academy of Sciences. 2014;111. https://doi.org/10.1073/pnas.1416737111

43. Wells A, Yates C, Shepard CR. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. Clin Exp Metastasis. 2008;25:621–8. https://doi.org/10.1007/s10585-008-9167-1

## Supplementary Data

1. Variant Prediction Data
https://docs.google.com/spreadsheets/d/1-YWGlG_QZtjaJDzVkKrXj5Vz1sdakTVQ/edit?usp=sharing&ouid=111301874115491980128&rtpof=true&sd=true

2. SIFT score table
https://docs.google.com/spreadsheets/d/1xRrUGs3yzOKdBOToxg5U6jJDHN0-Y2Rp/edit?usp=sharing&ouid=111301874115491980128&rtpof=true&sd=true

3. Polyphen Prediction
https://docs.google.com/spreadsheets/d/1uMe0joz5_6-vYp0mWiSMIOkGqF3mkW1-/edit?usp=sharing&ouid=111301874115491980128&rtpof=true&sd=true

4. Functional Study of Mutated protein
https://docs.google.com/spreadsheets/d/13pk1zl8YB-JEgT0qeWQI38eGV5QNG3Lp/edit?usp=sharing&ouid=111301874115491980128&rtpof=true&sd=true

**Abbreviation**

| | |
|---|---|
| CDH1 | *Cadherin-1* |
| SNP | Single Nucleotide Polymorphism |
| SIFT | Sorting Intolerant from Tolerant |
| POLYPHEN2 | Polymorphism Phenotyping v2 |
| CADD | Combined Annotation Dependent Depletion |
| EC | E-cadherin |
| HDGC | Hereditary Diffuse Gastric Cancer |
| EMT | Epithelial - Mesenchymal Transition |
| NGS | Next-Generation Sequencing |
| WGS | Whole Genome Sequencing |

| | |
|---|---|
| WES | Whole Exome Sequencing |
| WTS | Whole Transcriptome Sequencing |
| NCBI | National Center for Biotechnology Information |
| FASTA | Fast All |
| VCF | Variant Call Format |
| GRAVY | Grand Average of Hydropathy |
| PFAM | Protein families database |
| PRINTS | Protein Fingerprints database |
| PROSITE | Database of Protein Sites and patterns |
| SMART | Simple Modular Architecture Research Tool |
| PANTHER | Protein Analysis Through Evolutionary Relationships |
| OMIM | Online Mendelian Inheritance in Man |
| ACMG | American College of Medical Genetics and Genomics |