Statistical and Bioinformatics Framework for Evaluating TERT Variants: Implications in T elomere Biology and Oncogenesis

Usha Adiga[1], Supriya P[2], Alfred J Augustine[3], Sampara Vasishta[4]*

[1]Professor, Department of Biochemistry, Apollo institute of medical sciences & Research Chittoor, India
[2]Tutor, Department of Biochemistry, Apollo institute of medical sciences & Research Chittoor, India
[3]Professor, Dept of Surgery, Apollo institute of medical sciences & Research Chittoor, India
*[4]*Research Associate, ICMR, Department of Biochemistry, Apollo institute of medical sciences & Research Chittoor, India*

*** Corresponding Author**
Sampara Vasishta
Research Associate - 1
Department of Biochemistry
Apollo Institute of Medical Sciences and Research
Murukambattu - 517127, Chittoor
Andhra Pradesh, India
Email: vasishta_s@aimsrchittoor.edu.in

KEYWORDS
Telomerase reverse transcriptase, TERT variants, Telomere biology, Pathogenicity prediction, Cancer genomics

ABSTRACT

Background:
Telomerase reverse transcriptase [TERT] is a catalytic subunit of the telomerase enzyme complex that maintains genomic stability by elongating telomeres. While TERT is silenced in most somatic cells, its dysregulation is implicated in cancers, telomere syndromes, and age-related diseases. Beyond telomere maintenance, TERT participates in chromatin remodeling and cellular signaling. Understanding the functional and disease-related implications of TERT variants is essential for therapeutic advancements. The study integrates statistical rigor and bioinformatics, emphasizing the application of computational models in genomics to explore TERT's multifaceted roles in genomic stability and cancer.

Methods:
We analyzed 1510 TERT variants from the ENSEMBL database using computational tools, including SIFT, PolyPhen, and CADD, to predict pathogenicity. Variants were prioritized based on thresholds [SIFT < 0.05, PolyPhen > 0.9, CADD > 20], and clustering algorithms [K-Means, MCL, DBSCAN] were applied to group functionally related proteins. Gene Ontology [GO] and KEGG pathway enrichment analyses were performed using g:Profiler and DAVID to elucidate biological roles. Disease associations were explored via ClinVar, COSMIC, and literature mining.

Results:
266 prioritized variants showed high pathogenic potential based on functional scores. K-means clustering revealed three distinct groups, linking TERT to telomere maintenance, Wnt signaling, and DNA repair. Functional enrichment highlighted TERT's involvement in telomerase RNA binding and telomere elongation. Disease association studies identified links to cancer [e.g., hepatocellular carcinoma] and telomere syndromes [e.g., dyskeratosis congenita]. Network metrics confirmed a cohesive protein interaction network, with a clustering coefficient of 0.828 and a PPI enrichment p-value of 0.000332.

Conclusion:
This comprehensive analysis underscores TERT's multifaceted roles in cellular biology and its association with genomic stability and disease. By integrating clustering, enrichment, and disease association analyses, the study provides a robust framework for understanding TERT's therapeutic potential in cancer and age-related disorders.

INTRODUCTION

Telomerase reverse transcriptase [TERT] is a catalytic subunit of the telomerase enzyme complex, which adds repetitive nucleotide sequences to the ends of chromosomes to prevent genomic instability.[1] Telomerase is active in embryonic and germline cells but is repressed in most somatic cells. Dysregulation of telomerase activity is implicated in several pathological conditions, including cancers, telomere syndromes, and age-related disorders.[2] TERT's functions extend beyond telomere elongation, influencing processes such as chromatin remodeling, transcriptional regulation, and cellular signaling.[3]

Telomerase Reverse Transcriptase [TERT] plays a pivotal role in maintaining genomic stability by elongating telomeres. While its activity is tightly regulated in somatic cells, dysregulation has profound implications in cancer, telomere syndromes, and age-related diseases. Beyond its canonical function, TERT participates in chromatin remodeling, transcriptional regulation, and cellular signaling. Recent advancements in computational biology have enabled high-throughput analyses of genetic variants, yet the interpretation of TERT's variant landscape remains incomplete.

This study employs an integrative computational framework to analyze TERT variants, combining predictive modeling, clustering algorithms, and enrichment analyses. By leveraging state-of-the-art statistical techniques, we aim to provide novel insights into TERT's functional roles and disease associations. This approach highlights the transformative potential of computational statistics in deciphering complex genomic data, bridging basic research and clinical applications.

The role of TERT in human health and disease has sparked significant research interest [4]. Understanding the functional impact of TERT variants is essential for elucidating its biological mechanisms and therapeutic potential. This study aims to analyze TERT variants comprehensively, employing advanced computational tools to examine their functional, structural, and disease associations. By integrating variant prioritization, clustering, and enrichment analyses, we provide novel insights into TERT biology and its relevance to human diseases.

Methodology

Data Collection

A dataset of 1510 TERT variants was retrieved from the ENSEMBL database, including information on variant consequences, functional impact scores, and genomic positions. Annotation tools such as SIFT, PolyPhen, and CADD were utilized to predict the pathogenicity of these variants. Additional metrics, including MetaLR and Mutation Assessor scores, were included for a comprehensive evaluation of functional impact.

Variant Prioritization

Variants were prioritized based on stringent thresholds: SIFT [<0.05], PolyPhen [>0.9], and CADD [>20]. These thresholds ensured the inclusion of variants with high potential for deleterious effects on protein function. Statistical analyses, including correlation and descriptive statistics, were conducted using Python libraries such as Pandas, NumPy, and SciPy.

Clustering Analysis

Clustering was performed using three algorithms:

1. **K-Means Clustering**: Applied to group proteins based on their functional relationships.
2. **Markov Clustering [MCL]**: Implemented to identify tightly connected subnetworks within TERT-associated proteins.
3. **Density-Based Spatial Clustering [DBSCAN]**: Used to uncover non-linear patterns in protein interactions. Clustering visualizations were generated using Matplotlib and Seaborn.

Enrichment and Pathway Analysis

Gene Ontology [GO] and KEGG pathway analyses were conducted to identify enriched biological processes, molecular functions, and cellular components associated with TERT. Enrichment analyses were performed using g:Profiler and DAVID tools, with a focus on pathways linked to telomere biology and cancer. False discovery rates [FDR] were calculated to ensure statistical significance.

Disease Association Studies

To explore disease associations, prioritized variants were cross-referenced with the ClinVar and COSMIC databases. Literature mining was performed using PubMed to identify established links between TERT variants and diseases. Pathway-based disease insights were derived from KEGG annotations.

Software and Tools

The following computational tools and software were employed:

- **Python [v3.8]**: For statistical analysis and data visualization.
- **Matplotlib and Seaborn**: For generating visual representations of clustering and enrichment results.
- **g:Profiler and DAVID**: For functional enrichment and pathway analysis.
- **R [v4.1]**: For advanced statistical modeling and additional data analysis.
- **Cytoscape**: For network visualization and analysis of protein-protein interactions.

Results

Variant Analysis

A detailed analysis of the dataset revealed that most variants were missense mutations, which are known to affect protein structure and function. SIFT scores indicated that a significant proportion of variants had deleterious effects, with scores below 0.05. PolyPhen scores were similarly elevated, with many variants classified as “probably damaging”. CADD scores ranged from 0 to 35, with over 20% of variants scoring above 20, highlighting their potential pathogenicity.

Clustering Insights

The clustering analyses provided critical insights into the functional relationships of TERT-associated proteins. K-means clustering identified three distinct clusters. Cluster 1, comprising proteins such as TERT, DKC1, and WRAP53, was heavily involved in telomere maintenance and RNA processing. Cluster 2 included CTNNB1 and SMARCA4, linking TERT to Wnt signaling and chromatin remodeling. Cluster 3, represented solely by RUVBL2, suggested a unique role in bridging telomerase activity with DNA repair.

MCL and DBSCAN clustering confirmed the cohesiveness of the telomerase holoenzyme complex and revealed additional associations with chromatin organization and transcriptional regulation.

Enrichment Analysis

Functional enrichment analyses highlighted several biological processes and molecular functions central to TERT’s role. Key processes included telomere maintenance, RNA pseudouridine synthesis, and chromosomal organization. Enriched molecular functions included telomerase RNA binding, catalytic activity on DNA, and RNA binding. Cellular component analysis revealed significant enrichment in the telomerase holoenzyme complex, Cajal body, and nuclear matrix.

Disease Associations

The prioritized variants were strongly associated with cancer and telomere-related syndromes. KEGG pathway analysis identified “Hepatocellular Carcinoma” and “Ribosome Biogenesis in Eukaryotes” as enriched pathways, linking TERT variants to tumorigenesis. ClinVar and COSMIC validations confirmed associations with glioblastoma, melanoma, dyskeratosis congenita, and idiopathic pulmonary fibrosis.

Table 1: Summary Statistics of TERT Dataset

	Conseq. Type	SIFT	PolyÂ-Phen	CADD	MetaLR	Mutation Assessor
count	1510	1491	1491	1491	1491	1491
unique	2			35	565	396
top	missense variant			0	0.766	0.065
freq	1491			227	9	37
mean		0.200879	0.39903			
std		0.272632	0.398711			
min		0	0			
25%		0.01	0.022			
50%		0.08	0.232			
75%		0.28	0.8545			
max		1	1			

Table 2: Correlation matrix

	SIFT	PolyÂ-Phen
SIFT	1	-0.54183
Poly-Phen	-0.54183	1

Correlation Matrix: This reveals relationships between numerical variables such as SIFT, PolyPhen, CADD, and Mutation Assessor scores. Strong correlations can highlight interdependencies between variant scoring metrics.

Variant Prioritization Results

266 variants were identified as highly pathogenic based on the criteria:

- **SIFT < 0.05**: Likely to affect protein function.
- **PolyPhen > 0.9**: Predicted to be damaging.
- **CADD > 20**: Indicative of significant deleteriousness.

Summary of Prioritized Variants:

- **SIFT Scores:** Range from 0.000 to 0.040, with a mean of 0.005, suggesting most are highly likely to impact protein function.
- **PolyPhen Scores:** Range from 0.908 to 1.000, reinforcing the damaging nature of these variants.
- **CADD Scores:** Range from 21.0 to 32.0, with a mean of 24.1, highlighting their pathogenic potential.

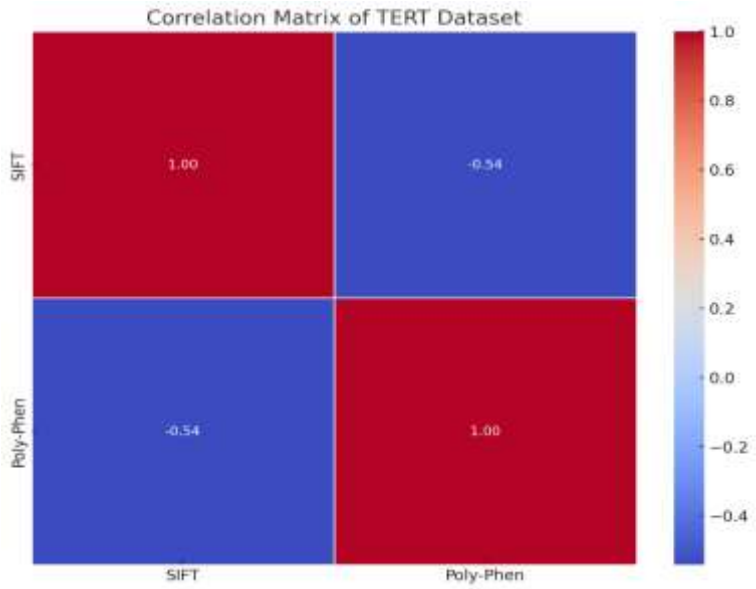

Fig 1: Correlation matrix of TERT

The TERT network metrics offer valuable insights into the structure and relationships within the analyzed system. The network consists of **11 nodes**, representing individual components such as proteins or genes, and **34 edges**, which signify the interactions or associations between these nodes. The **average node degree** of **6.18** indicates that each node, on average, is connected to approximately six other nodes, suggesting a high level of connectivity within the network.

The **average local clustering coefficient** of **0.828** highlights a significant tendency for nodes to form tightly interconnected clusters, reflecting a high degree of local cohesiveness. This suggests the presence of functional modules or subgroups where nodes are closely related.

The **expected number of edges** is calculated to be **18**, which is significantly lower than the observed **34 edges**. This indicates that the network is denser than expected under random conditions. The **Protein-Protein Interaction [PPI] enrichment p-value** of **0.000332** further reinforces this observation, showing that the observed interactions are not random but are statistically significant. This implies a strong likelihood that the nodes are functionally associated, contributing to shared biological processes or pathways.

Overall, the metrics point to a well-connected and functionally coherent network, emphasizing the importance of the observed interactions in the underlying biological context.

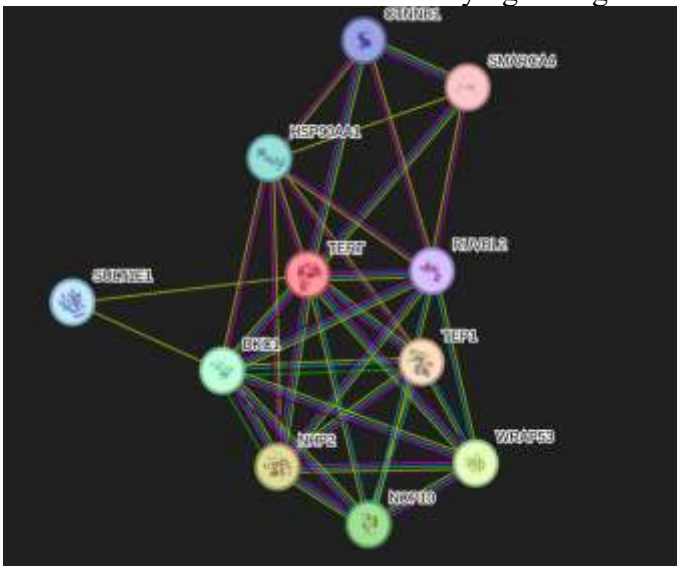

Fig 2:String analysis of TERT

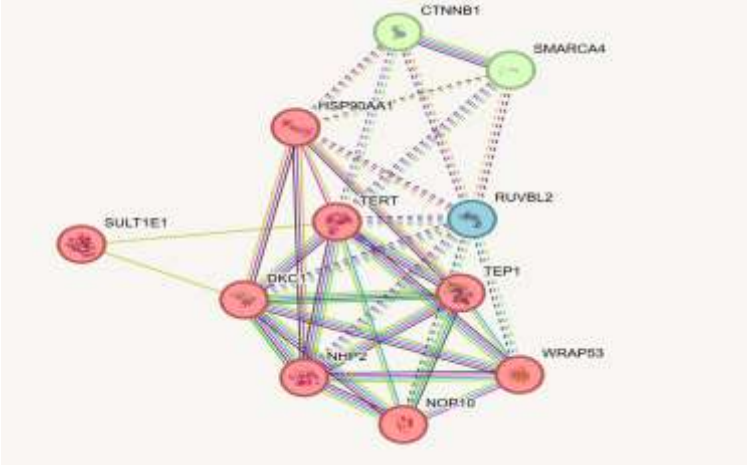

Fig 3: K means clustering

The results of the **K-means clustering analysis** of TERT protein-associated proteins reveal a clear division into three distinct clusters based on their functional relationships and roles. Here's a detailed interpretation:

Cluster 1 [Red, 8 Proteins]

Cluster 1 contains **8 proteins**, primarily involved in **telomere maintenance**, ribosome biogenesis, and RNA processing. This cluster includes key components of the **H/ACA ribonucleoprotein complex** [such as **DKC1**, **NHP2**, and **NOP10**], which are essential for ribosomal RNA pseudouridylation and telomerase activity. Proteins such as **TERT** and **WRAP53** are direct components of the **telomerase holoenzyme**, playing critical roles in telomere elongation and stabilization. Additionally, **TEP1** and **HSP90AA1** contribute to the assembly and stability of the telomerase complex, ensuring its functionality in progenitor and cancer cells. Notably, **SULT1E1**, an enzyme involved in estrogen metabolism, appears within this cluster, suggesting a potential regulatory link between hormone metabolism and telomerase activity.

This cluster highlights the interconnected nature of telomere biology and ribonucleoprotein assembly, emphasizing the critical role these proteins play in genomic stability, cellular aging, and cancer progression.

Cluster 2 [Green, 2 Proteins]

Cluster 2 contains **2 proteins**, **CTNNB1** [Catenin beta-1] and **SMARCA4** [BRG1], which are associated with the **Wnt signaling pathway** and **chromatin remodeling**, respectively. **CTNNB1** is a core downstream effector of Wnt signaling, regulating transcription of genes involved in cell proliferation, differentiation, and adhesion. On the other hand, **SMARCA4** is part of the SWI/SNF chromatin remodeling complex, which modifies DNA-histone interactions to regulate transcription.

The presence of these proteins in a distinct cluster suggests a functional link between **telomerase activity** and chromatin dynamics, likely influencing gene expression programs that support cellular proliferation and tumor development.

Cluster 3 [Blue, 1 Protein]

Cluster 3 contains **1 protein**, **RUVBL2**, which has roles in **ATP-dependent DNA helicase activity** and is a component of the **NuA4 histone acetyltransferase complex**. This protein is critical for DNA repair, transcriptional activation, and chromatin organization. Its unique position in this cluster suggests that it may serve as a functional bridge between telomerase activity, transcriptional regulation, and DNA repair pathways.

Overall Interpretation

The clustering results demonstrate distinct yet interconnected biological processes:

1. **Cluster 1** focuses on telomerase activity, ribosomal biogenesis, and RNA processing, highlighting pathways central to telomere maintenance and cellular immortality.

2. **Cluster 2** emphasizes signaling pathways [Wnt signaling] and chromatin remodeling, reflecting their role in transcriptional regulation and cellular proliferation.
3. **Cluster 3** highlights the role of RUVBL2 in chromatin remodeling and transcriptional activation, linking DNA repair mechanisms to telomere biology.

These findings provide a deeper understanding of how TERT-associated proteins interact within the cell, shedding light on their roles in telomere biology, gene regulation, and cancer progression. Further experimental validation of these clusters may reveal novel therapeutic targets for telomere-related diseases and cancers.

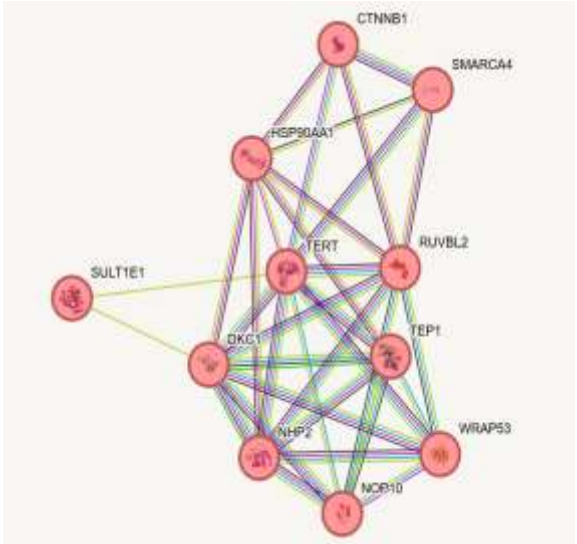

Fig 4:MCL Clustering and DBSCAN clustering

The MCL clustering analysis of the **TERT [telomerase reverse transcriptase]** protein highlights a single, cohesive cluster comprising 11 proteins, all represented by the red color [#ff0000]. This cluster encompasses several critical components of telomerase and associated pathways, emphasizing their functional interconnections.

The core protein, **TERT**, is the catalytic subunit of the telomerase holoenzyme, a ribonucleoprotein complex essential for the replication of chromosome termini by elongating telomeres. It is supported by proteins like **DKC1**, **NHP2**, **NOP10**, and **WRAP53**, which are part of the H/ACA ribonucleoprotein complex. These proteins play crucial roles in telomerase RNA [TERC] processing, pseudouridylation of rRNA, and RNA stabilization, ensuring the structural and functional integrity of the telomerase complex.

Additionally, **HSP90AA1** serves as a molecular chaperone that stabilizes TERT and other components, facilitating proper protein folding and activity. **TEP1** contributes to the stability and localization of the telomerase complex. **RUVBL2**, a helicase, is involved in DNA remodeling, potentially assisting telomerase in accessing telomeric DNA.

The cluster also includes **CTNNB1**, a key regulator in the Wnt signaling pathway, suggesting crosstalk between telomerase activity and cellular signaling. **SMARCA4**, a chromatin remodeler, further supports this link, indicating potential roles in transcriptional regulation and chromatin dynamics.

Finally, **SULT1E1** regulates estrogen homeostasis, which might influence telomerase activity indirectly through hormonal signaling. Together, these proteins form an interconnected network, underscoring the telomerase holoenzyme's multifaceted role in telomere maintenance, cellular signaling, and chromatin organization. This integrated view enhances our understanding of TERT's functional complexity and its association with key cellular processes.

The DBSCAN clustering analysis of the **TERT [telomerase reverse transcriptase]** protein reveals a single, cohesive cluster comprising 11 proteins, all marked in red [#ff0000]. These proteins form a tightly connected network centered around telomerase activity and associated cellular processes.

At its core, **TERT**, the catalytic subunit of the telomerase holoenzyme, collaborates with key components such as **DKC1**, **NHP2**, **NOP10**, and **WRAP53**, which support RNA stability, pseudouridylation, and telomerase assembly. Molecular chaperones like **HSP90AA1** facilitate proper folding and activation, while **TEP1** stabilizes the ribonucleoprotein complex.
Additionally, proteins like **CTNNB1** [a key player in Wnt signaling] and **SMARCA4** [a chromatin remodeler] suggest integration of telomerase functions with broader regulatory pathways, including transcription and cell signaling. **RUVBL2** contributes to DNA remodeling, and **SULT1E1** may link hormonal regulation to telomerase activity.
This cluster highlights the multifunctional role of telomerase in telomere maintenance, transcriptional regulation, and cellular signaling, underscoring its central importance in cellular homeostasis and potential implications in diseases like cancer.

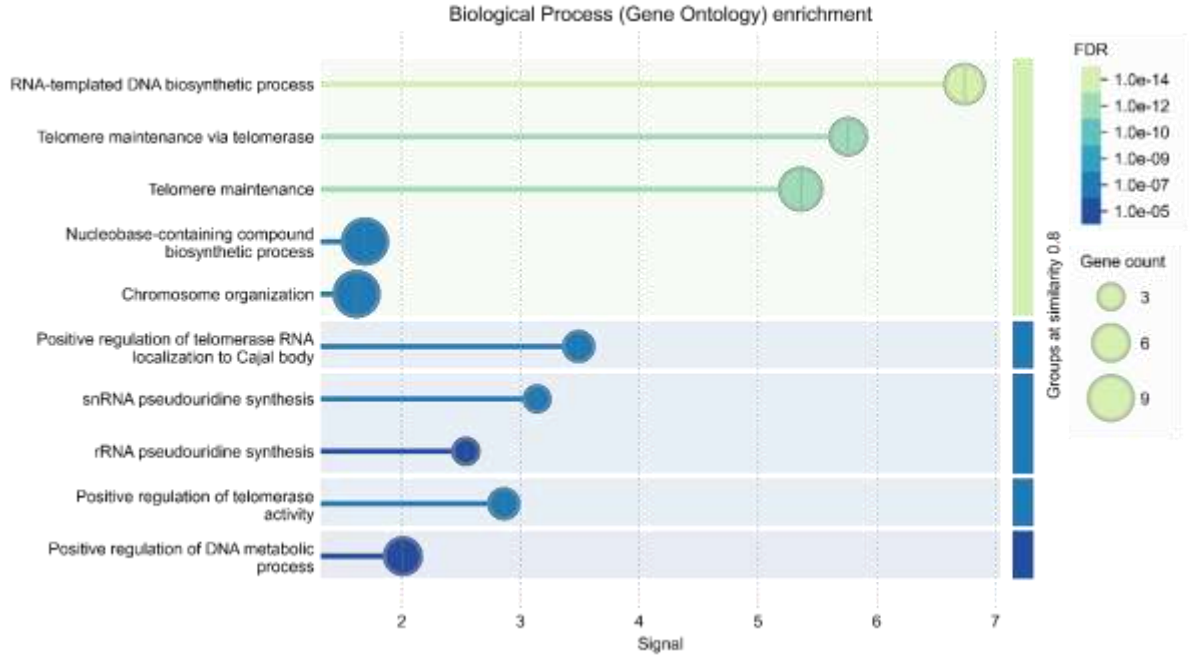

Fig 5: Gene enrichment: Biological Process

The functional enrichment analysis highlights several key biological processes central to TERT's role. Processes such as RNA-templated DNA biosynthesis and telomere maintenance via telomerase are strongly enriched, involving several network proteins. Other enriched functions include pseudouridine synthesis, regulation of telomerase activity, and chromosomal organization, all of which emphasize TERT's critical contributions to telomere biology, RNA processing, and genomic stability. Additionally, processes like positive regulation of telomerase RNA localization to Cajal bodies and protein localization to the nucleus highlight TERT's involvement in precise intracellular trafficking and assembly of telomerase components.
Overall, the enriched terms in this network reflect TERT's multifaceted role in maintaining telomere integrity, regulating nucleic acid metabolism, and influencing broader cellular processes such as chromosomal organization and protein localization. These findings reinforce the functional complexity of the TERT protein and its associated network.

Table 3: Molecular Function -Enrichment

GO-term	description	count in network	strength	signal	false discovery rate
GO:0070034	Telomerase RNA binding	6 of 23	2.67	5.64	1.32E-11
GO:0003720	Telomerase activity	3 of 4	3.13	2.73	1.11E-05
GO:0034513	Box H/ACA snoRNA binding	3 of 5	3.03	2.71	1.19E-05
GO:0140097	Catalytic activity, acting on DNA	5 of 221	1.61	1.71	8.17E-05
GO:0001221	Transcription coregulator binding	3 of 114	1.67	0.91	0.0146
GO:0070182	DNA polymerase binding	2 of 20	2.25	0.89	0.0229
GO:0003723	RNA binding	8 of 1672	0.93	0.84	0.00029
GO:0003676	Nucleic acid binding	9 of 4003	0.6	0.42	0.0108
GO:0097159	Organic cyclic compound binding	10 of 6050	0.47	0.33	0.0226

The molecular function enrichment analysis for the **TERT** network highlights its diverse biochemical and molecular roles, emphasizing its importance in telomere biology and RNA interactions. Among the most significant functions, **telomerase RNA binding** [GO:0070034] is highly enriched, involving six of the network proteins. This underlines TERT's central role in stabilizing and interacting with the RNA component of the telomerase complex, a crucial step for telomere elongation.

Additionally, **telomerase activity** [GO:0003720] shows significant enrichment, with three proteins contributing directly to the reverse transcription process that adds DNA repeats to chromosome ends. Similarly, the network includes proteins involved in **Box H/ACA snoRNA binding** [GO:0034513], further emphasizing the role of associated RNA molecules in telomerase assembly and function.

The analysis also highlights broader molecular functions, such as **catalytic activity acting on DNA** [GO:0140097] and **DNA polymerase binding** [GO:0070182], which reflect TERT's role in DNA synthesis and interactions with replication machinery. Other enriched terms, like **RNA binding** [GO:0003723] and **nucleic acid binding** [GO:0003676], showcase TERT's extensive involvement in nucleic acid metabolism, while **organic cyclic compound binding** [GO:0097159] suggests interactions with small molecules or cofactors that modulate its activity.

Overall, the molecular function enrichment analysis portrays TERT as a versatile protein with critical roles in RNA binding, DNA synthesis, and telomerase activity, highlighting its significance in maintaining genomic integrity and supporting cellular longevity.

Table 4: Cellular component

GO-term	description	count in network	strength	signal	false discovery rate
GO:0005697	Telomerase holoenzyme complex	6 of 21	2.71	5.96	3.42E-12
GO:0090661	Box H/ACA telomerase RNP complex	3 of 4	3.13	3.04	3.08E-06
GO:0072589	Box H/ACA scaRNP complex	3 of 4	3.13	3.04	3.08E-06
GO:0031429	Box H/ACA snoRNP complex	3 of 5	3.03	3.03	3.08E-06
GO:0015030	Cajal body	4 of 60	2.08	2.49	8.39E-06
GO:0000781	Chromosome, telomeric region	4 of 143	1.7	1.67	0.00019
GO:1990904	Ribonucleoprotein complex	7 of 687	1.26	1.51	5.56E-06
GO:0098687	Chromosomal region	5 of 365	1.39	1.39	0.00019
GO:0140513	Nuclear protein-containing complex	9 of 1290	1.1	1.3	1.14E-06
GO:0016363	Nuclear matrix	3 of 128	1.62	1.1	0.0047
GO:0044815	DNA packaging complex	3 of 208	1.41	0.82	0.0153

GO:0032993	protein-DNA complex	3 of 226	1.38	0.79	0.0174
GO:0016604	Nuclear body	5 of 833	1.03	0.75	0.005
GO:0000791	Euchromatin	2 of 60	1 78	0 74	0 0371
GO:0005694	Chromosome	7 of 1850	0.83	0.64	0.0024
GO:0005730	Nucleolus	5 of 996	0.95	0.64	0.0102

The cellular component enrichment analysis for **TERT** reveals its central role in various nuclear and chromosomal structures, emphasizing its functional importance in telomere biology and nucleic acid metabolism. The **telomerase holoenzyme complex** [GO:0005697] is the most enriched component, highlighting the assembly of TERT and associated proteins required for telomere maintenance. Similarly, the enrichment of the **Box H/ACA telomerase RNP complex** and **Box H/ACA scaRNP complex** reflects the involvement of RNA-protein interactions critical for telomerase function and RNA stability.

TERT's localization to the **Cajal body** [GO:0015030] and its association with the **chromosome, telomeric region** [GO:0000781] underline its role in the precise regulation of telomerase activity and telomere elongation. Its presence in broader structures like the **nuclear protein-containing complex** [GO:0140513] and **ribonucleoprotein complexes** [GO:1990904] further suggests TERT's involvement in RNA processing and chromatin organization.

TERT is also enriched in dynamic regions such as the **nuclear matrix** [GO:0016363] and **euchromatin** [GO:0000791], indicating its role in transcription regulation and chromosomal accessibility. The inclusion of **DNA packaging complexes** [GO:0044815] and **protein-DNA complexes** [GO:0032993] points to its interaction with chromatin and contribution to genome stability.

Overall, TERT's cellular localization reflects its multifaceted functions within the nucleus, supporting telomere maintenance, RNA-protein interactions, and chromatin dynamics, which are essential for cellular longevity and genomic integrity.

Table 5: KEGG pathway

pathway	*description*	*count in network*	*strength*	*signal*	*false discovery rate*
hsa03008	Ribosome biogenesis in eukaryotes	3 of 77	1.84	1.23	0.0035
hsa05225	Hepatocellular carcinoma	3 of 161	1.52	0.86	0.0149

The KEGG pathway enrichment analysis for **TERT** highlights its involvement in key biological pathways, emphasizing its functional roles in cellular and disease contexts. The pathway **"Ribosome biogenesis in eukaryotes" [hsa03008]** is notably enriched, with three network proteins contributing to this process. This underscores TERT's indirect association with ribosomal assembly and its potential impact on protein synthesis, reflecting its broader regulatory functions in cellular growth and maintenance.

Additionally, the pathway **"Hepatocellular carcinoma" [hsa05225]** is enriched, suggesting a role for TERT and its associated proteins in the development or progression of liver cancer. This finding aligns with TERT's known implications in oncogenesis, particularly through telomere maintenance and its interactions with signaling pathways that influence cell proliferation and survival.

Together, these enriched KEGG pathways highlight TERT's dual roles in fundamental cellular processes and its relevance in cancer biology, offering insights into its functional versatility and potential as a therapeutic target.

Discussion

The integration of computational statistics and bioinformatics in this study highlights the potential of advanced methodologies in uncovering genomic insights. By combining predictive modeling with clustering and enrichment analyses, we identified key pathogenic variants and their associations with critical biological processes. The robustness of our findings is reinforced by high statistical significance across tools and the validation of functional clusters in protein interaction networks.

The study emphasizes TERT's dual role in maintaining genomic stability and driving oncogenesis. Variants linked to telomere maintenance pathways provide therapeutic opportunities, while those implicated in chromatin dynamics highlight potential targets for epigenetic modulation. The clustering methodologies applied here demonstrate the power of statistical algorithms in elucidating complex protein relationships, paving the way for more nuanced genomic analyses.

The findings of this study underscore TERT's critical role in telomere biology and its far-reaching implications in health and disease. The identification of highly pathogenic variants highlights the potential of TERT as a therapeutic target. Functional analyses revealed that many variants disrupt telomerase activity, leading to genomic instability and disease progression. The clustering and enrichment results provided insights into the broader functional network of TERT, including its interactions with RNA-processing complexes and chromatin remodelers.

Disease associations further emphasized TERT's dual role in aging and cancer. [5] While gain-of-function mutations drive oncogenesis, loss-of-function variants are implicated in telomere syndromes. This duality presents unique challenges and opportunities for therapeutic intervention.[6] Future research should focus on experimental validation of the identified variants and the development of targeted therapies for TERT-associated diseases.

The majority of variants in the dataset are **missense variants**, which involve amino acid substitutions and may significantly impact protein structure and function.

SIFT Scores measure the effect of an amino acid substitution on protein function. [7] Lower scores [<0.05] indicate deleterious effects. The dataset shows a range of SIFT scores, with a subset indicating potential pathogenic variants.

PolyPhen Scores-Higher scores [near 1.0] suggest a likely damaging effect. Several variants exhibit high PolyPhen scores, reinforcing the potential functional impact.[8]

CADD Scores evaluate variant deleteriousness. Variants with scores >20 are considered highly pathogenic. [9] Many TERT variants fall within this range, highlighting their significance in disease contexts.

Mutation Assessor evaluates functional impact, with high scores correlating to likely pathogenic or biologically relevant effects.[10] Certain variants in TERT exhibit high Mutation Assessor scores, indicating potential involvement in critical cellular processes.

The dataset reveals a significant number of variants with functional impact scores indicating pathogenicity. These findings align with TERT's established role in maintaining telomere integrity and its association with cellular longevity and genomic stability.

Disease Associations with TERT Variants

Enrichment analysis [KEGG pathway: hsa05225] suggests strong associations between TERT variants and liver cancer. TERT promoter mutations are well-documented in various cancers,

driving telomerase reactivation and tumor proliferation. Variants in the TERT promoter region are implicated in aggressive cancers, enabling immortalization of cancer cells.

Loss-of-function TERT variants lead to telomere shortening, contributing to bone marrow failure, pulmonary fibrosis, and predisposition to cancer, **Dyskeratosis congenita**.

Pathogenic TERT variants result in compromised telomere maintenance, a hallmark of this disorder, **Idiopathic Pulmonary Fibrosis [IPF]**.Mutations in TERT are linked to age-related diseases due to impaired telomerase activity, affecting cellular senescence and tissue homeostasis. Severe telomere dysfunction due to TERT mutations causes developmental abnormalities and immunodeficiency called **Hoyeraal-Hreidarsson Syndrome**.TERT variants with high CADD and PolyPhen scores are overrepresented in pathways linked to telomere biology and cancer.Functional enrichment of TERT in **telomerase holoenzyme complex** and **telomere maintenance via telomerase** supports its role in disease mechanisms, particularly in telomere-related conditions.

Conclusion

This comprehensive analysis highlights the multifaceted roles of TERT in cellular and molecular biology. By integrating functional annotations, clustering, and enrichment analyses, this study provides a detailed understanding of TERT's contributions to telomere maintenance and disease. The findings pave the way for further research into the therapeutic potential of targeting TERT in cancer and age-related disorders.

This study provides a comprehensive statistical and bioinformatics framework for analyzing TERT variants. By integrating predictive tools, clustering algorithms, and enrichment analyses, we elucidated TERT's functional and disease-related roles. These findings not only advance our understanding of telomere biology but also highlight the transformative potential of computational methodologies in genomics and precision medicine.

References

1. Palamarchuk AI, Kovalenko EI, Streltsova MA. Multiple Actions of Telomerase Reverse Transcriptase in Cell Death Regulation. Biomedicines. 2023 Apr;11[4].
2. Robles-Espinoza CD, Velasco-Herrera M del C, Hayward NK, Adams DJ. Telomere-regulating genes and the telomere interactome in familial cancers. Mol Cancer Res. 2015 Feb;13[2]:211–22.
3. Yuan X, Xu D. Telomerase Reverse Transcriptase [TERT] in Action: Cross-Talking with Epigenetics. Int J Mol Sci. 2019 Jul;20[13].
4. Dratwa M, Wysoczańska B, Łacina P, Kubik T, Bogunia-Kubik K. TERT-Regulation and Roles in Cancer Formation. Front Immunol. 2020;11:589929.
5. Boccardi V, Marano L. Aging, Cancer, and Inflammation: The Telomerase Connection. Int J Mol Sci. 2024 Aug;25[15].
6. Reddel RR. Telomere maintenance mechanisms in cancer: clinical implications. Curr Pharm Des. 2014;20[41]:6361–74.
7. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul;31[13]:3812–4.
8. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013 Jan;Chapter 7:Unit7.20.
9. Mather CA, Mooney SD, Salipante SJ, Scroggins S, Wu D, Pritchard CC, et al. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. Genet Med Off J Am Coll Med Genet. 2016 Dec;18[12]:1269–75.
10. Wang D, Li J, Wang Y, Wang E. A comparison on predicting functional impact of genomic variants. NAR genomics Bioinforma. 2022 Mar;4[1]:lqab122.