

Integrated Analysis of Gene Expression, Protein Structure, and Evolutionary Relationships: A Case Study Using GEPIA, PDB, and Phylogenetic Tools

Swati Singh

PhD Scholar, Department of Biotechnology, Sharda University, Greater Noida, India

Pankaj Taneja

Professor, Department of Biotechnology, Sharda School Bioscience and Technology (SSBT) Sharda University, Greater Noida, India

Abstract

Lymphomas represent a highly heterogeneous group of hematological malignancies arising from lymphoid tissues and characterized by diverse clinical behaviors, molecular signatures, and genetic aberrations. Their pathogenesis is driven by a complex interplay of genomic alterations, epigenetic dysregulation, and disrupted signaling pathways that collectively promote uncontrolled cellular proliferation, impaired apoptosis, and immune evasion. In the present study, we conduct an integrative bioinformatics investigation focusing on four key genes strongly implicated in lymphomagenesis—CD30 (TNFRSF8), ALK (Anaplastic Lymphoma Kinase), NPM1 (Nucleophosmin), and TP53 (Tumor Protein p53). These genes are known to regulate essential biological processes such as cell-cycle control, chromatin organization, protein transport, and oncogenic signaling cascades. To explore their transcriptional dynamics, gene expression profiles were systematically analyzed using GEPIA (Gene Expression Profiling Interactive Analysis), integrating RNA-seq data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) database. Comparative expression assessments across various lymphoma subtypes and corresponding normal tissues revealed distinct expression patterns, highlighting potential subtype-specific oncogenic roles and diagnostic relevance. Structural characteristics of the encoded proteins were examined through experimentally validated data retrieved from the Protein Data Bank (PDB). Domain architecture, ligand-binding regions, and conformational stability were evaluated to understand their functional mechanisms and mutation-induced structural alterations that contribute to malignant transformation. Emphasis was placed on the oncogenic ALK fusion variants, the multifunctional scaffold protein NPM1, the tumor suppressor TP53, and the immunoregulatory receptor CD30, each demonstrating unique structural features that underpin their pathogenic influence. Furthermore, the evolutionary conservation and divergence of these genes were explored by constructing phylogenetic trees using protein sequences from a broad range of eukaryotic organisms. The phylogenetic analysis provided insights into their evolutionary trajectories, revealing conserved functional domains essential for survival and lineage-specific variations that may correlate with differential susceptibility to lymphoid malignancies across species. Collectively, this integrative analysis offers a comprehensive understanding of the molecular architecture, oncogenic potential, and evolutionary context of CD30, ALK, NPM1, and TP53. The study not only broadens current knowledge of lymphoma biology but also establishes a robust foundation for the development of targeted therapeutic strategies, precision diagnostics, and future experimental investigations aimed at functional validation.

Materials & Methods:

Gene expression was analyzed using GEPIA.
Protein structures were studied using PDB.
Phylogenetic trees were constructed using UniProt sequences.

Results:

CD30 and ALK were highly expressed in lymphoma.
NPM1 showed stable expression.
TP53 showed altered expression in aggressive cases.
Protein structures revealed important functional domains.

Discussion :

CD30 and ALK act as major therapeutic targets.
NPM1 and TP53 maintain genome stability and apoptosis.
Structural and expression data support their role in lymphoma.