

Feb 24, 2026

Genotyping and computational analysis of variants rs200161705 and rs772747361 in *NEK1* and variants rs748112833 and rs142030898 in *TBK1*

DOI

<https://dx.doi.org/10.17504/protocols.io.3byl48q68vo5/v1>

Diolina Gonçalves da Silva¹, Nayane Soares de Lima¹, Caroline Christine Pincela da Costa¹,
Stela Silva Nolêto Leite de Carvalho¹, Rodrigo da Silva Santos^{1,2}, Angela Adamski da Silva Reis^{1,2}

¹Núcleo de Pesquisas em Neurogenética (NeuroGene), Instituto de Ciências Biológicas (ICB II), Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brazil;

²Departamento de Bioquímica e Biologia Molecular, Instituto de Ciências Biológicas (ICB II), Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brazil.



diolinagoncalves

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.3byl48q68vo5/v1>

Protocol Citation: Diolina Gonçalves da Silva, Nayane Soares de Lima, Caroline Christine Pincela da Costa, Stela Silva Nolêto Leite de Carvalho, Rodrigo da Silva Santos, Angela Adamski da Silva Reis 2026. Genotyping and computational analysis of variants rs200161705 and rs772747361 in NEK1 and variants rs748112833 and rs142030898 in TBK1. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.3byl48q68vo5/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: January 19, 2026

Last Modified: February 24, 2026

Protocol Integer ID: 238865

Keywords: computational analysis of variants rs200161705, nek1 variant, rs772747361 in nek1, tbk1 variant, genome stability, variants rs748112833, tbk1 this protocol details taqman qpcr genotyping, als research, protocol details taqman qpcr genotyping, nek1, variants rs200161705, innate immune signaling, rs748112833

Abstract

This protocol details TaqMan qPCR genotyping of *NEK1* variants (rs200161705, rs772747361) and *TBK1* variants (rs748112833, rs142030898), followed by *in silico* functional and structural analyses (SIFT, PolyPhen-2, PANTHER, NetSurfP-2.0, ConSurf, SWISS-MODEL, DynaMut, SCooP). Applications include ALS research focusing on DNA repair, autophagy/innate immune signaling, and genome stability.

Before start

Be sure to wear a lab coat, mask, and gloves.

Handle all reagents in a clean, dedicated pre-PCR area, and be especially careful when manipulating every component of the master mix to prevent contamination.

Protect TaqMan® probes from light at all times, as VIC® and FAM™ dyes are photosensitive; keep tubes covered, minimize exposure to ambient light, and store aliquots appropriately to preserve fluorescence stability.

Study design and ethics

- 1 In this protocol, the genotyping of *NEK1* and *TBK1* variants was performed using quantitative polymerase chain reaction (qPCR), followed by computational analysis conducted with widely used online bioinformatic tools. All methodological steps were described to ensure the reliability, reproducibility, and validity of the study. The research was approved by the Ethics and Research Committee of the Federal University of Goiás (CEP/UFG), Brazil (CAAE 79593117.7.0000.5083). All participants were recruited voluntarily and provided written informed consent prior to enrollment.

Variant selection

- 2 Four single-nucleotide variants (SNVs) were selected for this protocol. Two variants, rs200161705 and rs772747361, are located in the *NEK1* gene (HGNC: 7744; NCBI Gene: 4750; Ensembl: ENSG00000137601; OMIM: 604588; UniProt: Q96PY6), positioned on chromosome 4q33 [1]. The other two variants, rs748112833 and rs142030898, are located in the *TBK1* gene (HGNC: 11584; NCBI Gene: 29110; Ensembl: ENSG00000183735; OMIM: 604834; UniProt: Q9UHD2), positioned on chromosome 12q14.2 [2]. These SNVs were selected based on prior evidence of involvement in DNA-damage response, autophagy pathways, and neurodegeneration.

Molecular analysis

- 3 Peripheral blood (10 mL) was collected in EDTA tubes, labeled, and stored at 4 °C until processing. Genomic DNA was extracted using the PureLink™ Genomic DNA kit (Invitrogen/Thermo Fisher®) and stored at -20 °C. DNA purity and concentration were assessed using a NanoDrop™ ND-1000 spectrophotometer, targeting an A260/280 ratio of approximately 1.8–2.0. All samples were then diluted to a working concentration of 10 ng/μL for use in the qPCR genotyping assays. For the molecular analysis, TaqMan™ hydrolysis probes were used on a QuantStudio™ 6 Pro Real-Time PCR System. Each reaction contained genomic DNA (10 ng/μL), TaqMan™ Genotyping Master Mix (2x), and pre-designed probes from the Applied Biosystems™ TaqMan™ SNP Genotyping Assays.

Table 1. Target sequence corresponding to the TaqMan hydrolysis probe for *NEK1* and *TBK1* genes.

	A	B	C	D
	Gene	Variant	Context Sequence [VIC/FAM]	Assay ID

A	B	C	D
	(ID)		
NE K1	rs77274 7361	TTTAAATAACTGAGACACC AACTG[C/T]GGAGATCAT AGGAATAATGCAAAGA	C_38650501 9_10
NE K1	rs20016 1705	CTGAGGAGAGAGAACTTT TCAATG[C/T]GTTTGGCTA TAAACCTTTCTCCAA	C_19133356 6_10
TB K1	rs74811 2833	TATTAGTATGAAGAAATTAA AGGAA[G/A]AGATGGAAGG GGTGGTTAAAGAACT	C_3252386 49_10
TB K1	rs14203 0898	TCAGGAATTAATGCGAAAG GGGATA[C/T]GATGGCTGA TGTAAGTAATAGATTG	C_16189302 2_10

Thermo Fisher Scientific® , 2017 [3].

The probe has two distinct fluorophores: VIC and FAM, responsible for fluorescence when the corresponding allele is identified in the target sequence. One allele is labeled with the fluorescent dye VIC, designed to detect allele 1 (allele wild) in the corresponding target sequence, while the second allele is labeled with the dye FAM and targets allele 2 (allele mutant).

The TaqMan™ Genotyping Master Mix is supplied ready to use. To this solution, a hydrolysis probe specific to the TaqMan™ SNP Genotyping Assays (Applied Biosystems™ – Thermo Fisher Scientific®), provided at 20X, was added. Following the manufacturer's recommendation for overage, a 10% safety margin was applied. Thus, the reagent–probe mixture was prepared for 106 reactions (corresponding to a 96-well plate plus 10% excess). For each reaction well, 9 µL of reagents were combined with 1 µL of genomic DNA (pre-diluted to 10 ng/µL), resulting a final volume of 10 µL per well. Reagents were prepared and pre-diluted samples were dispensed in a laminar flow cabinet, following contamination-control best practices and the equipment manufacturer's protocol. After loading all wells, the plate was sealed with MicroAmp™ Optical 96-Well adhesive film.

Note (recommended controls): Include no-template controls (NTCs) to monitor contamination and, when available, known genotype controls to verify clustering.

The thermal cycling reactions were performed according to the standard parameters recommended by the manufacturer for TaqMan™ hydrolysis probe-based SNP

genotyping assays [4]. All reactions were executed in Standard Mode on the QuantStudio™ 6 Pro Real-Time PCR System. The cycling profile is shown in Table 2 and includes an initial enzyme activation step, followed by denaturation and allele-specific annealing/extension phases. These conditions ensure optimal probe hybridization, polymerase activity, and fluorescent signal discrimination for genotype calling.

Table 2. Cycling protocol using validated genotyping assays for qPCR in QuantStudio 6 Pro.

A	B	C	D
Stage	Temperature	Duration	Cycles
Enzyme activation	95 °C	10 minutes	Hold
Denaturation	95°C	15 seconds	Hold
Anneal/Extension	60 °C	1 minute	40

Thermo Fisher Scientific® , 2017 [3].

3.1 Interpretation of results

Genotyping data were processed using the Diomni™ Design and Analysis (RUO) platform within the Thermo Fisher Cloud Genotyping module. The software generated automated scatter plots that allowed visualization and classification of allelic signals. Data interpretation was performed through two complementary analytical outputs: amplification curves and allelic discrimination plots.

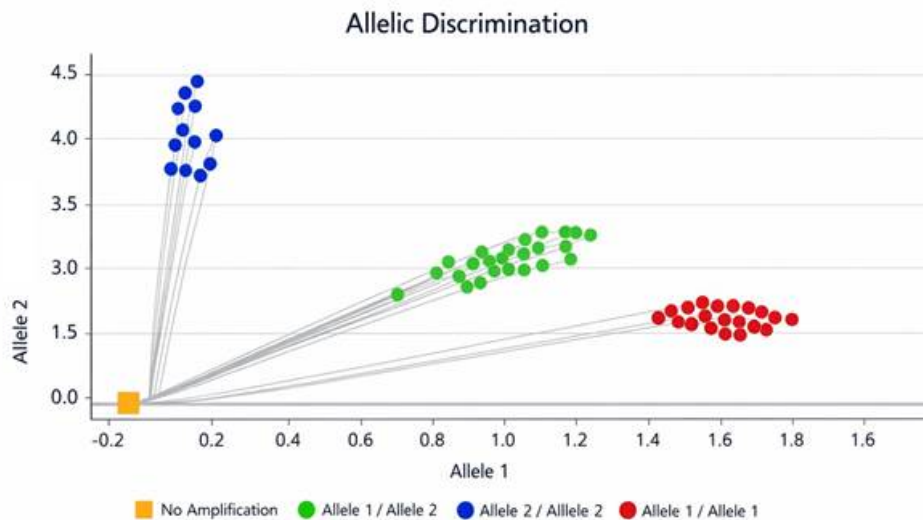
The amplification curves display fluorescence intensity as a function of cycle number, enabling verification of amplification kinetics and confirmation of probe-specific signal generation for each sample. These curves ensure that only reactions exhibiting valid exponential amplification are included in the genotype call.

The allelic discrimination plot displays the endpoint fluorescence intensities for Allele 1 (x-axis) and Allele 2 (y-axis), allowing assignment of each sample to its corresponding genotype cluster. Each point represents an individual reaction, and its position in the two-dimensional fluorescence space reflects the relative signal generated by the allele-specific VIC and FAM probes.

Samples with no amplification appear near the origin and are shown in yellow.

The heterozygous group (Allele 1 / Allele 2) forms an intermediate cluster with balanced fluorescence levels, represented in green. Samples mutant homozygous for Allele 2 (Allele 2 / Allele 2) display high Allele 2 signal and low Allele 1 signal, forming the upper-left cluster, shown in blue. Conversely, samples wild homozygous for Allele 1 (Allele 1 / Allele 1) exhibit strong Allele 1 fluorescence and minimal Allele 2 signal, forming the lower-right cluster, shown in red. Clear spatial separation among clusters demonstrates robust probe specificity, reliable fluorescence discrimination, and high-quality endpoint data for confident genotype calling.

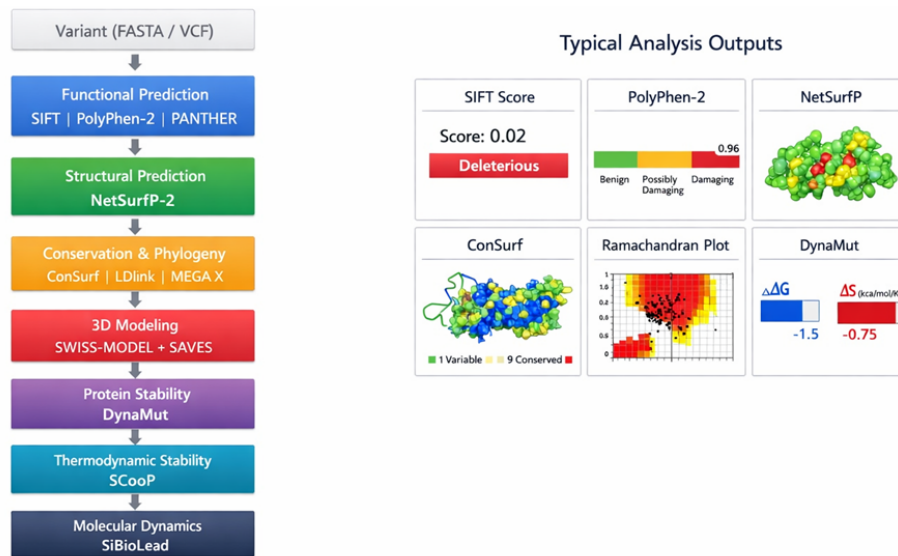
Figure 1. Allelic discrimination plot generated from TaqMan® SNP genotyping reactions.



In silico analysis

- 4 To complement wet-lab genotyping, we conducted an integrated *in silico* assessment of NEK1 and TBK1 variants spanning functional prediction, structural context, evolutionary conservation, three-dimensional modeling, stability, docking, and molecular dynamics. This multi-tool workflow (SIFT, PolyPhen-2, PANTHER, NetSurfP-2.0, ConSurf/LDlink/MEGA X, SWISS-MODEL/SAVES, DynaMut, SCOP, HADDOCK, and SiBioLead) provides orthogonal evidence to prioritize variants and to rationalize their potential molecular impact on protein structure, interaction networks, and conformational dynamics (Figure 2).

Figure 2. Workflow for *in silico* functional, structural, and evolutionary analysis of *NEK1* and *TBK1* variants and examples of typical computational outputs.



Left: Sequential pipeline used for functional, structural, evolutionary, stability, docking, and molecular dynamics analyses of *NEK1* and *TBK1* variants. Right: Example outputs generated by each tool, including SIFT and PolyPhen-2 scores, NetSurfP surface accessibility, ConSurf conservation map, Ramachandran plot, and DynaMut stability prediction

Prediction of functional impact

- 5 The potential functional effects of *NEK1* and *TBK1* variants were assessed using three complementary *in silico* tools: Sorting Intolerant From Tolerant (SIFT), Prediction of Functional Effects of Human nsSNPs (PolyPhen-2), and Protein Analysis Through Evolutionary Relationships (PANTHER). All platforms require as input the wild-type amino acid sequence in FASTA format and the specific amino acid substitution for each variant.

Interpretation of functional prediction scores

- PolyPhen-2 scores range from 0 to 1; variants with values >0.85 are considered damaging [4].
- SIFT also ranges from 0 to 1; substitutions with scores <0.05 are predicted to be deleterious [5].

- PANTHER provides the Position-Specific Evolutionary Preservation (PSEP) index, representing evolutionary preservation time; highly conserved sites are more likely to be functionally critical [6].

Prediction of functional impact

- 6 NetSurfP-2.0 (<https://services.healthtech.dtu.dk/services/NetSurfP-2.0>) [7] was employed to assess the structural consequences of *NEK1* and *TBK1* variants. This tool predicts residue-specific features such as Relative Surface Accessibility (RSA) and Accessible Surface Area (ASA), providing insight into whether substitutions occur in exposed or buried regions of the protein and how they may alter local structural environments.

Analysis of conservation profile, phylogeny and haplotypes

- 7 Evolutionary conservation analysis was carried out using ConSurf software (https://consurf.tau.ac.il/consurf_index.php) [8], which assigns each position a conservation score from 1 (variable) to 9 (highly conserved), based on Bayesian and maximum-likelihood phylogenetic modeling. Haplotype block structure and allele frequencies were examined using Ldlink (<http://analysistools.nci.nih.gov/LDlink/>) [9], while the evolutionary history of *NEK1* and *TBK1* proteins was reconstructed using MEGA X to generate phylogenetic trees. This information helps in identifying protein residues and regions critical for specific functions.

Protein modeling and model quality assessment

- 8 Protein modeling and quality assessment three-dimensional structures of the wild-type and mutant proteins were modeled using SWISS-MODEL server (<https://swissmodel.expasy.org/>), based on FASTA input sequences [10]. Model quality was evaluated using the SAVES platform (<https://saves.mbi.ucla.edu/>), which produces stereochemical validation metrics and a Ramachandran plot, where: 90% residues in favored regions = high-quality model 0% residues in disallowed regions = ideal geometry. Ramachandran plot, favorable regions are shown in red, additional allowed regions in dark yellow, generously allowed regions in light yellow, and disallowed regions in white. Black and red dots indicate the positions of the protein's amino acids.

Prediction of interatomic interactions

- 9 The server outputs the predicted change in stability (kcal/mol), along with the entropy energy variation between the wild-type and mutant structures (kcal/mol/K). Prediction of protein stability as a function of temperature. the predicted change in Gibbs free energy ($\Delta\Delta G$, in kcal/mol), which reflects the stabilizing or destabilizing effect of the

mutation, and the variation in vibrational entropy (ΔS , in kcal/mol/K), which indicates changes in molecular flexibility. Negative $\Delta\Delta G$ values correspond to destabilizing mutations, whereas positive values indicate increased structural stability.

The server outputs the predicted change in stability (kcal/mol), along with the entropy energy variation between the wild-type and mutant structures (kcal/mol/K). Prediction of protein stability as a function of temperature.

Prediction of protein stability as a function of temperature

- 10 SCooP predictor (<http://babylone.3bio.ulb.ac.be/SCooP/index.php>) [12] was employed to derive the complete thermodynamic stability profile of each protein variant using the corresponding structural models as input. The server generates full stability curves and reports key thermodynamic parameters, including the melting temperature (T_m), which reflects the temperature at which half of the protein population unfolds; the heat capacity change upon folding (ΔC_p), an indicator of the energetic cost associated with structural compaction; and the folding enthalpy at T_m (ΔH_m), which quantifies the heat absorbed or released during the unfolding transition. Together, these metrics provide a temperature-dependent view of protein stability, complementing static stability predictions obtained from other computational tools.

Docking

- 11 Molecular docking analyses were performed for NEK1-ATP and TBK1-OPTN complexes in both wild-type and mutant forms using the HADDOCK 2.4 server (<https://rascar.science.uu.nl/haddock2.4/>) [13]. HADDOCK integrates biochemical, biophysical, and bioinformatic information to guide the docking process through the definition of *active* and *passive* residues, allowing the generation of biologically meaningful interaction models. All parameters were kept at default settings except for the sampling configuration, which was adjusted to 10.000 / 400 / 400 for the it0 / it1 / water refinement stages, respectively, due to the limited availability of experimental restraints.

Docked conformations were ranked according to the HADDOCK Score (HS), a weighted combination of Van der Waals, electrostatic, desolvation, and restraint violation energies. Lower HS values correspond to more energetically favorable and structurally plausible interaction models.

Molecular dynamics simulation

- 12 To investigate the dynamic and structural behavior of *NEK1* and *TBK1* variants over time, molecular dynamics (MD) simulations were performed for both wild-type and mutant proteins using the SiBioLead MD Simulation Server (<https://sibiolead.com/MDSIM>). All systems were parameterized with the OPLS/AA force field and placed in a triclinic simulation box, followed by solvation using the

SPC/E explicit water model. System neutrality was achieved by adding 0.15 M NaCl. Energy minimization was carried out via the steepest descent algorithm for 5.000 steps. Systems were equilibrated under NVT and NPT ensembles at 300 K and 1 bar, respectively. Production simulations were run for 100 ns using the Leap-Frog integrator, recording 1.000 frames per trajectory. Trajectory data were evaluated using standard MD metrics, including:

- Root Mean Square Deviation (RMSD) – global structural stability
- Root Mean Square Fluctuation (RMSF) – local residue flexibility
- Radius of Gyration (Rg) – overall compactness
- Solvent Accessible Surface Area (SASA) – exposure to solvent
- Hydrogen bond profiles – stability of intermolecular and intramolecular interactions

These parameters collectively describe the conformational stability, mobility, and dynamic behavior of each protein variant under physiological conditions.

Protocol references

- 1-Chatterjee, N.; Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, v. 58, n. 5, p. 235–263, 2017. <https://doi.org/10.1002/em.22087>
- 2-Konopka, A.; Atkin, J.D. DNA Damage, Defective DNA Repair, and Neurodegeneration in Amyotrophic Lateral Sclerosis. *Frontiers in Aging Neuroscience*, v. 14, 27 abr. 2022. PAUL, S. et al. Disrupted autophagy overactivates TBK1 and results in mitotic defects promoting chromosomal instability. *Autophagy*, 16 jan. 2026. <https://doi.org/10.3389/fnagi.2022.786420>
- 3-Thermo Fisher Scientific (2025). QuantStudio 6 and 7 Pro Real-Time PCR Systems. <https://www.thermofisher.com/br/en/home/life-science/pcr/real-time-pcr/real-time-pcr-instruments/quantstudio-systems/models/quantstudio-6-7-pro.html>
- 4-Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. et al [Current protocols in human genetics], v. Chapter 7, n. 1, p. Unit7.20, 2013. <https://doi.org/10.1002/0471142905.hg0720s76>
- 5-Sim, N.L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research*, v. 40, n. Web Server issue, p. W452–7, 2012. <https://doi.org/10.1093/nar/gks539>
- 6-Thomas, P.D. et al. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein science: a publication of the Protein Society*, v. 31, n. 1, p. 8–22, 2022. <https://doi.org/10.1002/pro.4218>
- 7-Klausen, M.S. et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins*, v. 87, n. 6, p. 520–527, 2019. <https://doi.org/10.1002/prot.25674>
- 8-Ashkenazy, H. et al. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, v. 38, n. Web Server issue, p. W529–533, 1 jul. 2010. <https://doi.org/10.1093/nar/gkq399>
- 9-Machiela, M.J.; Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants: Fig.1. *Bioinformatics*, v. 31, n. 21, p. 3555–3557, 2 jul. 2015. <https://doi.org/10.1093/bioinformatics/btv402>
- 10-Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, v. 46, n. W1, p. W296–W303, 2018. <https://doi.org/10.1093/nar/gky427>
- 11-Rodrigues, C.H.; Pires, D.E.; Ascher, D.B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*, v. 46, n. W1, p. W350–W355, 2018. <https://doi.org/10.1093/nar/gky300>
- 12-Pucci, F.; Kwasigroch, J. M.; Rooman, M. SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics (Oxford, England)*, v. 33, n. 21, p. 3415–3422, 2017. <https://doi.org/10.1093/bioinformatics/btx417>
- 13-Honorato, R. V. et al. The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature protocols*, 2024. <https://doi.org/10.1038/s41596-024-01011-0>.