Bioinformatics Analysis Identifies NDRG1 Gene Variants that may be Clinically Relevant

Borré, GB1, Pimenta AT1, Chmieleski GS1, Moyses GR1, Sousa SCB1, Rabi LT1, Peres KC1, Teixeira ES1, Bufalo NE1,2,3, Ward LS1*

Affiliation:
1Laboratory of Cancer Molecular Genetics, School of Medical Sciences, University of Campinas (UNICAMP), Campinas/SP, Brazil
2Department of Medicine, São Leopoldo Mandic and Research Center, Campinas/SP, Brazil
3Department of Medicine, Max Planck University Center, Campinas/SP, Brazil

*Corresponding author:
Laura Sterian Ward, Laboratory of Cancer Molecular Genetics, School of Medical Sciences, PO Box 6111, Campinas, SP-Brazil

Citation: Borré GB, Pimenta AT, Chmieleski GS, Moyses GR, Sousa SCB, Rabi LT, Peres KC, Teixeira ES, Bufalo NE, Ward LS. Bioinformatics Analysis Identifies NDRG1 Gene Variants that may be Clinically Relevant. Journal of Bioinformatics and Systems Biology 5 (2022): 158-162.

Received: November 28, 2022
Accepted: December 05, 2022
Published: December 14, 2022

Abstract

Background: The search of single nucleotide variants that might have the capacity to alter genetic information and influence in regular cellular pathways, enhancing expansion, mitosis and evasion capacity to neoplasm cells, is central in understanding the molecular nature of distinct cellular growth abnormalities and is critical because it might expose new possibilities for therapeutic targets. The expression of NDRG1 protein, encoded by NDRG1 gene, has already been correlated with tumor progression and evasion, but information on different types of neoplasm is still contentious.

Objective: To explore probable correlations of susceptibility, progression and clinical characteristics between NDRG1 gene polymorphisms (SNPs) and patients that developed thyroid tumors.

Methods: SNPs were obtained from the NCBI dbSNP. The encoded protein primary sequences were got from the UniProt database. We employed the three FASTA primary sequences to analyze the amino acid changes. The bioinformatics tools used were: PredictSNP1.0 (which encompasses: PANTHER, SNAP, PolyPhen-1, PhD-SNP, nsSNPAnalyze, SIFT, PredictSNP, PolyPhen-2, MAPP,); I-Mutant2.0; MUpro; PROVEAN; Haploview and SNPs3D).

Results: The NCBI database reports 319 missense SNPs in the NDRG1 gene. The SIFT tool predicted that 51 nsSNPs of 109 (which means 46.78%) were deleterious; the SNAP tool predicted nearly 30%; PolyPhen-2, 53 (48.62%); 52 (47.70%) derived from PhD-SNP; PolyPhen-1 indicated 38 nsSNPs (approximately 35%); and MAPP showed 47 (which is 43%). Finally, the PredictSNP tool contemplated 13 (approximately 12%) nsSNPs deleterious by all integrated tools, including rs201348291 and rs15132213, whose scores were the most significant, thus indicating a higher possibility that these SNPs are correlated and influence the pathophysiology of thyroid neoplasm.

Conclusions: We demonstrated that NDRG1 rs201348291 and rs151322132 may be involved in thyroid cancer emergence and deserve further validation and evaluation of their clinical applicability in determining the risk of thyroid nodules malignancy and thyroid cancer prognostic.

Keywords: NDRG1; Thyroid; Bioinformatics; In silico analysis.

Introduction

Background: The majority of the modifications in human genetic material are called Single Nucleotide Polymorphisms (SNPs) and occur at a
single nucleotide base in DNA (1). These variations can be responsible for mutations and modify proteins regarding its function, structure and stability, therefore resulting in different disorders. SNPs can be useful in clinical practice as markers of diseases. They can help identify whether someone has a higher risk of developing a determined disease, also predicting its possible course, the patient’s response to medications, and therefore his prognosis. SNPs can help identify the individual response to certain agents such as environmental toxins, in addition to drugs, individualizing a therapeutic response profile and assisting in patient management (3). Nonsynonymous SNPs - nsSNPs or missense variations might alter the translation and transcription process and modify the amino acid residues sequence (4). Missense variations can destabilize and modify protein solubility and therefore cause malfunction and gene deregulation (4).

The NDRG1 protein is encoded by the NDRG1 gene, which responds to cellular stress signals (6, 7). NDRG1 participates in important cell signaling pathways and is associated with suppression of cell growth and differentiation (8,10,13). However, its role in tumor development and evolution is still controversial. It might have an anti-metastatic effect in some tumors, but, on the contrary, it has been correlated with more undifferentiated tumor, less benign prognosis and higher potential to more aggressiveness and metastasis (10, 11). In colorectal, breast and prostate neoplasms, the expression of NDRG1 presents with lower incidence of metastasis and progression (8,10). In contrast, it correlates with more aggressive prognosis in hepatocarcinoma (12,13). NDRG1 protein is likely to be involved in tumor development and evasion mechanisms through MMP-9 and MMP-2 whereas its effect on the metastatic potential appears to be associated with EMT and E-cadherin activities and also with aberrant methylation (8).

There is evidence of higher expression of the protein in thyroid carcinomas, primary and metastatic, in comparison to normal and benign thyroid affections in studies considering immunohistochemical (10). In patients with thyroid carcinoma its expression was correlated with an higher-risk AMES category and also more advanced TNM stages, suggesting it has a role in thyroid tumor progression (10). Furthermore, p.G136D is a novel carcinoma somatic alteration of NDRG1 that was found in a patient with PTC (33). An exome sequencing study proposed that variants of NDRG1 might have an important role in papillary thyroid cancer (PTC) development and can be possibly used as markers of higher aggressiveness and as prognostic indicators (31). Nonetheless, this gene's role in thyroid pathology is still very little understood.

This study aimed to investigate nsSNPs of the NDRG1 gene that could produce deleterious changes in the protein and could be associated with the carcinogenic process of the thyroid.

Materials and Methods

Data Selection

The SNPs for this study were selected from the NCBI database (https://www.ncbi.nlm.nih.gov/snp/?term=NDRG1). All three found FASTA sequences, three primary sequences of the proteins encoded by the gene, were contemplated for further analysis, which were taken from the UniProt database (https://www.uniprot.org/uniprot/Q92597) under the code Q92597.

Data analysis

The previously described bioinformatics instruments were used to decode and express the effects that the NDRG1 protein suffered from the DNA changes. These effects were analyzed through the following tools:

Deleterious nsSNPs prediction

PredictSNP1.0 (http://loschmidt.chemi.muni.cz/predictsnp1/) (15) was used to predict the impact on protein function and structure. This tool allows access to nine top performing prediction tools: PolyPhen-2, SIFT, PhD-SNP, MAPP, PolyPhen-1, PredictSNP and SNAP. PolyPhen-1 (Polymorphism Phenotyping) evaluates the possibility of impact of amino acid substitutions and PolyPhen-2 (Polymorphism Phenotyping v2) estimates the potential effect of an amino acid substitution on the structure and function of a human protein. SIFT (Sorting Intolerant from Tolerant) analyzes if an amino acid substitution modifies protein function (16). SNAP (Screening for Unacceptable Polymorphisms) is a neural network-based method used to predict functional effects of nsSNPs using information from in silico-derived proteins (19) MAPP (Multivariate Analysis of Protein Polymorphisms) investigates the physicochemical variation present in each column of a protein sequence alignment and predicts the impact of amino acid substitutions on protein function (17). PhD-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphisms) classifies nsSNPs that change the human DNA causing disease (18). PredictSNP1.0 shows the confidence scores generated by each tool and a consensus prediction as percentages (15).

Prediction of linkage disequilibrium

Pairs of altered alleles contained in genetic information can be more likely to be biologically inherited than other alleles from other pairs of SNPs. This is linkage equilibrium. The Haploview software informs the linkage disequilibrium prediction of the human genome (20).

Protein-protein interaction analysis

STRING (https://string-db.org/cgi/about) predicts and informs protein-protein interactions. These interactions can be physical or functional.
Probability analysis of the impact of nsSNPs on the function of the studied protein

The SNPs3D tool (http://www.SNPs3D.org) classifies and relates which genes can be considered to be involved in specific diseases. The tool also informs whether the nsSNPs modify protein function. The information derives from the analysis of the protein structure and stability and also of sequence conservation.

Results

SNP dataset

Data on the NDRG1 gene SNPs (NCBI Gene ID: 10397) were retrieved during September 2021 from the NCBI database (https://www.ncbi.nlm.nih.gov/snp/?term=ndrg1), from where 16,383 SNPs were obtained. The 319 that were missense were further evaluated.

Prediction of deleterious nsSNPs

147 out of the 319 missense polymorphisms obtained from dbSNP showed amino acid changes and some even had more than one change, so the three FASTA primary sequences were used to analyze these amino acid variations. They correspond to the three protein isoforms, are encoded by the gene in question and were taken from the UniProt database under the code Q92597.

SIFT showed that 51 out of 109 (46.78%) SNPs were deleterious; SNAP indicated 32 (29.35%) deleterious nsSNPs; PolyPhen-2 indicated 53 (48.62%) nsSNPs; PhD-SNP indicated 52 (47.70%) nsSNPs; PolyPhen-1 identified 38 (34.86%) nsSNPs; and MAPP indicated 47 (43.11%) deleterious nsSNPs. The PredictSNP tool analysis considered 13 (11.92%) nsSNPs deleterious by all the integrated tools, as shown in table 1.

Four SNPs (rs2272646, rs3779941, rs3088599 and rs2977497) from intronic regions found in literature were also analyzed, but were found to be mostly neutral.

Extensive literature search was performed for each of the deleterious nsSNPs selected by these tools, but we were unable to find any description or mention of any of them connected to thyroid nodules or thyroid tumors. Unfortunately, all these SNPs had very minor allele frequency (MAF < 0.01).

Analysis of Protein-Protein Interaction

The STRING tool (available at: https://string-db.org/cgi/network?taskId=b022LqIKmSV&sessionId=bxrG5RS2HYSO), provided results represented in figure 1. The server indicated that NDRG1 interacts with 11 proteins, including p53 cell tumor antigen (TP53), N-myc proto-oncogene protein (MYCN), serine/threonine-protein kinase (SGK1), NDRG2 protein (NDRG2), Myc proto-oncogene protein (MYC), WNT1-inducible signaling pathway protein 1 (WISP1), RAC-alpha serine/threonine-protein kinase (AKT1), and cadherin 1 (CDH1).

Discussion

The NDRG1 gene encodes a protein present in the cytosol and expressed in epithelial cells. While there is evidence that its activation has anti-metastatic action, other studies suggest that the protein expression is related to less tumor differentiation, more aggressive metastatic potential and, therefore, worse prognosis (10, 11). In fact, NDRG1 gene expression in breast (21, 28), liver (21, 22, 23, 25), lung (26) and cervical cancer (24) was positively correlated with disease recurrence and considered an indicator of poor prognosis for patient survival (26). On the other hand, the expression of NDRG1 in prostate (30), colon (29) and esophagus tumors (28) was associated with a favorable clinical evolution.

<table>
<thead>
<tr>
<th>nsSNPs</th>
<th>AA Change</th>
<th>FASTA</th>
<th>PredictSNP</th>
<th>SIFT</th>
<th>PhD-SNP</th>
<th>PolyPhen-1</th>
<th>PolyPhen-2</th>
<th>MAPP</th>
<th>SNAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs142426003</td>
<td>Y62C</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs2233319</td>
<td>M67V</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs201348291</td>
<td>P298S</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs368404338</td>
<td>R322C</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs144714216</td>
<td>R343H</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs366854363</td>
<td>Y7H</td>
<td>2</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs373637595</td>
<td>A21T</td>
<td>2</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs201348291</td>
<td>P232S</td>
<td>2</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs144714216</td>
<td>R277H</td>
<td>2</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs201759485</td>
<td>V189A</td>
<td>3</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs201348291</td>
<td>P217S</td>
<td>3</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs144714216</td>
<td>R262H</td>
<td>3</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>rs146613168</td>
<td>R262C</td>
<td>3</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>
An interesting strategy to investigate the role of a specific gene in the pathophysiology of a disease is the evaluation of the impact of this gene’s polymorphisms on the risk to develop this disease and/or its outcomes. Unfortunately, more than 300 missense polymorphisms have been reported in the human NDRG1 gene, and that number is still increasing. Not all of these polymorphisms include amino acid changes, and not all changes significantly affect the structure or function of the corresponding protein.

Data on the influence that nsSNPs of the NDRG1 gene have in the development of thyroid diseases, and their association with thyroid nodule clinical presentation and outcome have not yet been described in the literature. In order to select those that have the greatest potential to cause deleterious effects and influence the aetiopathogenesis of thyroid cancer, we undertook a thorough bioinformatics analysis of NDRG1 SNPs. We demonstrated that rs201348291 and rs151322132, two SNPs still poorly acknowledged, are potential markers of malignancy and may be related to the characteristics of thyroid cancer.

Conclusion

Our analysis indicated that these two nsSNPs deserve further validation, but our data have an obvious limitation, as we only performed in silico analyses. Clinical large-scale studies in different ethnic populations and laboratory experiments may provide more robust validation of our results. On the other hand, we provided solid foundation in the selection of SNPs of potential utility in the risk, evolution and prognosis of thyroid cancer patients.

Acknowledgments

The authors thank the group of the Laboratory of Cancer Molecular Genetics (GEMOCA) of the School of Medical Sciences (FCM-UNICAMP) for technical support. This study received financial support from the São Paulo Research Foundation (FAPESP grant 2020/16557-8) and INCT-TeraNano (CNPq nº 403193/2022-2). GBB received a research support from FAPESP. LSW is a Category 1 Research Fellow at the National Council for Scientific and Technological Development (CNPq).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References


