The role of PI3K signaling pathway and its associated genes in papillary thyroid cancer

Elham Amjad†, Solmaz Asnaashari† and Babak Sokouti*

Abstract

Background: One of the well-differentiated types of thyroid cancer is papillary thyroid cancer (PTC), often developed by genetic mutations and radiation.

Methods: In this study, the public NCBI-GEO database was systematically searched. The eligible datasets were the targets for biomarker discovery associated with PI3K signaling pathway.

Results: Only two datasets were suitable and passed the inclusion criteria. The meta-analysis outcomes revealed eleven upregulation and thirteen downregulation genes differentially expressed between PTC and healthy tissues. Moreover, the outcomes for survival and disease-free rates for each gene were illustrated.

Conclusions: The present research suggests a panel signature of 24 gene biomarkers in diagnosing the PTC.

Keywords: Papillary thyroid cancer, Meta-analysis, GEO datasets, Signaling pathway, Biomarker

Background

Papillary thyroid cancer (PTC), as one of the most prevalent cancer types diagnosed in the thyroid gland, covers more than 85% of total thyroid cancers [1, 2]. According to the fact that most of the large followed up thyroid cancer studies, specifically PTC, are epidemiologically carried out in the US nations, there are very few data available for monitoring the incidence of PTC worldwide [2]. Moreover, the well-known Surveillance, Epidemiology, and End Results database (SEER13) showed that the 5-year relative survival rate for 2010-2016 in the USA was 98.3% through age-adjusted incident rate with an unchanged trend for mortality rate of 0.4-0.5% [3]. Additionally, the Korean literature report indicated that the recurrence rate of patients with PTC after surgical treatment was less than 15% which led to survival by about 65% without including either 5- or 10-year survival rates [4]. However, various aggressive descriptors determined for PTC disease from imaging techniques have proposed a weak prognosis [4]. A recent case study reported that even if the diagnosis of PTC is somewhat tricky, combining a set of imaging techniques and fine-needle aspiration would make the road simplifying the PTC diagnosis [5].

The recent advancements in medical sciences have performed well by using microarray technology in narrowing down the list of diseases for diagnosis and prognosis procedures [6, 7], and on the other hand, the advances in technology will generate the Big data. In this regard, studying molecular and cellular functions of the particular diseases, e.g., the PTC, at genome-wide levels needs to be of priority. In other words, to differentiate between healthy and unhealthy tissues, the discovery of robust biomarkers is essential, either through experimental and clinical studies or computational biology approaches on Omics datasets [8]. There are several studies on investigating the molecular and physiological mechanisms of activation and inactivation of different signaling pathways having critical roles in the progress of PTC [9]. Since signaling pathways constitute hundreds of components, such as genes responsible for many cellular processes, their biological functions remain unclear and complicated [9].
Understanding the complex nature of PTC in terms of the involved signaling pathways urges studying vital roles of associated genes in the PI3K signaling pathway between PTC and healthy patients. For this purpose, the systematic search of the NCBI-GEO database retrieves the results of interest for further analysis and content screening based on inclusion and exclusion criteria. Finally, the genes of the selected GEO datasets will be the targets for identifying the differentially expressed genes among the datasets to be considered as robust biomarkers for PTC disease.

Methods
Identification of microarray datasets
The National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) database (i.e., http://www.ncbi.nlm.nih.gov/geo) was the repository source for the systematic search of microarray datasets. For this search, a Boolean query was used, including the keywords “papillary thyroid cancer” or “PTC.” Furthermore, a thorough inspection on the search results was necessary for identification and inclusion of those GEO datasets in the analysis which satisfy the following criteria: (i) based on “expression profiling by array,” (ii) to be of [9606] organism, (iii) mRNA sample types, (iv) to have both PTC samples and healthy controls, and (v) extracted from the source of thyroid tumor. Notably, any platform types were of interest. Moreover, the excluded GEO datasets were those that not fulfilled the abovementioned inclusion criteria. Accordingly, the final selected GEO datasets with sufficient data were prone to perform a meta-analysis by including the associated genes of the PI3K signaling pathway.

Associated genes for PI3K signaling pathway
The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (i.e., https://www.genome.jp/kegg/pathway.html) was the resource to derive the involved genes in the PI3K signaling pathway (hsa04151) for Homo sapiens organism. Then, to carry out the meta-analysis approach among the GEO datasets, the associated genes of the PI3K signaling pathway were only considered for this purpose.

Meta-analysis procedure
The ExAtlas free online tool for the meta-analysis of gene expression datasets possesses several main functions, such as the standard meta-analysis with fixed and random effects, z-score, and Fisher’s methods [10]. For this purpose, the input of the ExAtlas website included all GEO datasets meeting the inclusion criteria. However, among the whole gene symbols available in the GEO datasets, only those genes associated with the PI3K signaling pathway were selected for meta-analysis. The pre-processing stage of the input gene expression datasets comprised log2 transformation and quantile normalization applied to their corresponding intensity values as well as t-test ANOVA analysis. After, a data quality check of the included samples was carried out based on the standard deviation criterion SD ≤ 0.3. Finally, the false discovery rate (FDR) and fold change parameters were 0.05 and 2, respectively, for the meta-analysis stage. Due to the heterogeneous nature of the gene expression datasets, the results for the random-effects model would be necessary.

Analyses of survival and relapse-free rates
In survival analysis, various statistical methodologies analyzed the experimental data of interest in a defined period of follow-up time, usually, 200 months, in which death or relapse could happen carried out by plotting the Kaplan-Meier estimates [11, 12]. In this study, both overall survival (OS) and relapse-free survival (RFS) rates were considered on the Cancer Genome Atlas (TCGA)-THCA database professionally developed for thyroid carcinoma (n = 512) against control samples (n = 59). The Kaplan-Meier plots were obtainable using the GEPIA2 (http://gepia2.cancer-pku.cn/#index) web service [13]. In the KM-plot analysis, the p-values less than 0.05 were significant for input genes and the preset confidence interval for hazard ratio was 95%.

Results
The overall results obtained from systematically screening of the NCBI-GEO database, where the data extraction procedure is depicted is Fig. 1, by taking in to account the inclusion and exclusion criteria showed that only two microarray datasets (i.e., GSE29265: 20 Normal and 20 PTC and GSE97001: 4 Normal and 4 PTC) were eligible for meta-analysis procedure. Three hundred fifty-four out of a total of 354 genes involved in the PI3K signaling pathway were selected from the two GEO datasets to identify the significant genes differentially expressed between the two tissue types. The ExAtlas website demonstrated that all of the samples included in the GEO datasets passed the initial quality control, which then was suitable for the meta-analysis process based on a random-effects model due to the existence of possible heterogeneity. The meta-analysis revealed twenty-four genes significantly expressed in terms of p-value and FDR parameter between healthy and PTC samples, among which the numbers of upregulated and downregulated genes were eleven and thirteen, respectively, as represented in Fig. 2.
The overall survival (OS) and disease-free rates (RFS) for the identified upregulated and downregulated genes are illustrated in Figs. 3 and 4, respectively. Among upregulated genes, CCND2 was the only significant gene in terms of OS rate with p-value 0.017. Moreover, by considering the downregulated genes, three genes (i.e., GHR \( p \)-value=0.0035, FGF7 \( p \)-value=0.014, PRKCA \( p \)-value=0.045) were found to be significant in terms of OS rate; however, four genes (i.e., KIT \( p \)-value=0.012, GHR \( p \)-value=0.016, PGF \( p \)-value=0.05, FGFR2 \( p \)-value=0.029) were significant in terms of RFS rate.

**Discussion**

Despite the excellent reports on the prognosis of the papillary thyroid carcinoma (PTC) as the predominant form of thyroid cancer, estimating the overall survival of PTC patients has still been remained unknown [15]. In the current research, a meta-analysis approach could demonstrate the significant differentially expressed genes between two GEO datasets meeting the inclusion criteria with FDR<0.05. Avoiding any possible inconsistency between the datasets is critical such that only the GEO datasets with the source of PTC tissues were eligible. In total, 24 genes (11 upregulated and 13 downregulated) were differentially expressed in PTC patients while being compared to the healthy samples with the statistical significance of FDR <0.05 and \( p \)-value <0.05 (Table 1).

The above-listed genes in Table 1 were thoroughly inspected for their confirmation through various experimental studies considering the neoplastic thyroid disease. The expressions of associated genes in the PI3K signaling pathway were half downregulated and half upregulated. Taking in to account that some studies have reported on the target overexpressed gene whereas in
the current study was determined as downregulated, and vice versa, the main reasons for this may generally originate from several points such as viral infections, patients clinical history, treatment status, the source of control samples, the age of patients as well as patients’ race to mention a few (e.g., upregulation and downregulation of has-mir-345 in pancreatic cancer [44, 45]). Moreover, in the current meta-analysis study, statistical stages, including data normalization, t test, and ANOVA tests were performed on GEO datasets to compare the PTC and healthy tissues in the same conditions. Due to the error-prone nature of the clinical and experimental trials, several biases (e.g., publication, laboratory, environmental, and user biases) may affect the reported outcomes by the researchers, and hence, the gene expression levels may not be comparable [46]. The determined genes (with FC>2) in the current meta-analysis study can be useful in identifying potent biomarkers for future drug design and discovery. Among the identified biomarkers associated with the PI3K signaling pathway, four of them with FC > 4 were LAMB3 (upregulated), COMP (upregulated), KIT (downregulated), and PDGFRA (downregulated). Various studies have also confirmed the vital role of the activation of the PI3K signaling pathway in the progression and development of PTC disease [47, 48]. As described in the “Results” section, seven genes were significant while considering the OS and RFS rates; however, this outcome will not decline the fact that the other remaining biomarkers have vital roles in the development of the PTC disease.

Consequently, in the current meta-analysis research, a total of twenty-four genes associated with the PI3K signaling pathway were identified and thoroughly screened and validated via the experimental literature studies that could propose a panel of potential biomarkers in PTC disease.

Conclusion

The present study conducted on PTC GEO datasets revealed the significant role of the meta-analysis approach in determining the potential biomarkers for the disease. Eleven upregulated and thirteen downregulated genes were identified and validated through the literature investigations. By performing a meta-analysis study, one may conclude this type of analysis can fill the gaps between the computational and experimental studies; however, due to the existence of possible heterogeneities among the datasets, some of the differentially expressed genes may be missed that may urge novel algorithms to
Fig. 3 The survival analyses of identified upregulated genes in terms of OS and RFS
Fig. 4 The survival analyses of identified downregulated genes in terms of OS and RFS
cover the shortcomings. The biomarker discovery is one of the hot topics in the field, which still needs more advancements in terms of technical, experimental, and computational designs to achieve more robust and reliable biomarkers, and hence, to provide its vital role in diagnosis, prognosis, and treatment of diseases.

**Table 1** The list of 24 genes consistently expressed differentially between two GEO datasets associated with the PI3K signaling pathways in PTC patients seen in the literature

| Gene symbol | Chromosome (source: https://www.genecards.org/) | Disease (source: https://www.genecards.org/) | References for associated genes and PTC |
|-------------|-------------------------------------------------|-------------------------------------------------|----------------------------------------|
| KIT         | 4q12                                            | Gastrointestinal stromal tumors (GISTs), melanomas, lung cancer, and other tumor types | [16, 17]                               |
| PDGFRA      | 4q12                                            | Idiopathic hypereosinophilic syndrome, somatic and familial gastrointestinal stromal tumors, and a variety of other cancers | [18, 19]                               |
| IGF2        | 11p15.5                                         | One of the cancer-related genes                  | [20]                                   |
| GHR         | 5p13.1-p12                                      | Laron syndrome and growth hormone insensitivity, partial | [21]                                   |
| BCL2        | 18q21.33                                        | High grade B cell lymphoma with Myc and/or Bcl2 and/or Bcl6 rearrangement and follicular lymphoma 1 | [22]                                   |
| IRS1        | 2q36.3                                          | Diabetes mellitus, noninsulin-dependent and rare diabetes mellitus type 2 | [23]                                   |
| LPAR1       | 9q31.3                                          | Pulmonary fibrosis and spinal stenosis           | [24]                                   |
| FGF7        | 15q21.12                                        | Mucositis and acanthoma                          | [25]                                   |
| PGF         | 14q24.3                                         | Placental insufficiency and twin-to-twin transfusion syndrome | [26]                                   |
| FGFR2       | 10q26.13                                        | Lung and breast cancers                           | [27]                                   |
| PRKCA       | 17q24.2                                         | Chordoid glioma and papillary glioneuronal tumors | [28]                                   |
| LAMA2       | 6q22.33                                         | Muscular dystrophy, congenital mesosin-deficient, 1A and muscular dystrophy, limb-girdle, autosomal recessive 23 | [29]                                   |
| MYC         | 8q24.21                                         | Burkitt lymphoma and high grade B cell lymphoma with Myc and/or Bcl2 and/or Bcl6 rearrangement | [30]                                   |
| LAMB3       | 1q32.2                                          | Epidermolysis bullosa, junctional, Herlitz type, and epidermolysis bullosa, junctional, non-Herlitz type | [31]                                   |
| COMP        | 19p13.11                                        | Pseudoachondroplasia and epiphyseal dysplasia, multiple, 1 | [32]                                   |
| SPP1        | 4q22.1                                          | Pediatric systemic lupus erythematosus, and papillary cystadenocarcinoma | [33]                                   |
| TNC         | 9q33.1                                          | Deafness, autosomal dominant S6, and autosomal dominant non-syndromic sensori-neural deafness type Dfna | [34]                                   |
| ERBB3       | 12q13.2                                         | Lethal congenital contracture syndrome 2 and erythroleukemia, familial | [35, 36]                               |
| CCND1       | 11q13.3                                         | Von Hippel-Lindau syndrome and myeloma, multiple | [37]                                   |
| TLR2        | 4q31.3                                          | Leprosy 3 and colorectal cancer                  | [38]                                   |
| CCND2       | 12p13.32                                        | Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 3 | [39]                                   |
| LAMC2       | 1q25.3                                          | Epidermolysis bullosa, junctional, Herlitz type, and epidermolysis bullosa | [40]                                   |
| CDKN1A      | 6p21.2                                          | Multiple endocrine neoplasia, type I, and tongue cancer | [41]                                   |
| COL1A1      | 17q21.33                                        | Caffey disease and osteogenesis imperfecta, type I | [42, 43]                               |

**Authors’ contributions**
EA: Collected the data, review of datasets contents, performed the analysis, and draft manuscript preparation. SA: Study design, performed the analysis and discussion, draft manuscript preparation. BS: Conceived and designed the analysis, performed the analysis and discussion, draft manuscript preparation. The authors read and approved the final manuscript.

**Funding**
Not applicable

**Availability of data and materials**
Not applicable

**Declarations**

**Ethics approval and consent to participate**
Not applicable
40. Zhu W, Li C, Ai Z. Candidate agents for papillary thyroid cancer identified by gene expression analysis. Pathol Oncol Res. 2013;19(3):597–604. https://doi.org/10.1007/s12253-013-9625-1.

41. Yamashita AS, Geraldo MV, Fuziwara CS, Kulcsar MAV, Friguglietti CUM, da Costa RB, et al. Notch pathway is activated by MAPK signaling and influences papillary thyroid cancer proliferation. Transl Oncol. 2013;6(2):197. https://doi.org/10.1593/to.12442.

42. Lee K-Y, Huang SM, Li S, Kim J-M. Identification of differentially expressed genes in papillary thyroid cancers. Yonsei Med J. 2009;50(1):60–7. https://doi.org/10.3349/ymj.2009.50.1.60.

43. Liang W, Sun F. Identification of key genes of papillary thyroid cancer using integrated bioinformatics analysis. J Endocrinol Investig. 2018;41(10):1237–45. https://doi.org/10.1007/s40618-018-0859-3.

44. Lee EJ, Gusev Y, Jiang J, Nuovo GJ, Lerner MR, Frankel WL, et al. Expression profiling identifies microRNA signature in pancreatic cancer. Int J Cancer. 2007;120(5):1046–54. https://doi.org/10.1002/ijc.22394.

45. Lv Y, Huang S. Role of non-coding RNA in pancreatic cancer. Oncol Lett. 2019;18(4):3963–73. https://doi.org/10.3892/ol.2019.10758.

46. Boedigheimer MJ, Wolfinger RD, Bass MB, Bushel PR, Chou JW, Cooper M, et al. Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. BMC Genomics. 2008;9(1):285. https://doi.org/10.1186/1471-2164-9-285.

47. Shinohara M, Chung YJ, Saji M, Ringel MD. AKT in thyroid tumorigenesis and progression. Endocrinology. 2007;148(3):942–7. https://doi.org/10.1210/en.2006-0937.

48. Xu Y, Han YF, Zhu S-J, Dong J-D, Ye B. miRNA-148a inhibits cell growth of papillary thyroid cancer through STAT3 and PI3K/AKT signaling pathways. Oncol Rep. 2017;38(5):3085–93. https://doi.org/10.3892/or.2017.5947.

**Publisher’s Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.