Meta-analysis Identifies Endothelin-3 as a Prognostic Biomarker in Gastrointestinal Stromal Tumors

Zhiwei Qiao1, Fusako Kito2 and Tadashi Kondo1*

1Division of Rare Cancer Research, National Cancer Center Research Institute, Tokyo, Japan
2Department of Innovative Seeds Evaluation, National Cancer Center Research Institute, Tokyo, Japan

Abstract

Prognostic modalities have long been essential for selecting adjuvant treatment in patients with Gastrointestinal Stromal Tumor (GIST). The anatomic site of origin is a prognostic factor for GIST; GISTs originating in the small intestine (I-GISTs) have a considerably worse prognosis than GISTs originating in the stomach (S-GISTs). By studying the molecular backgrounds of GISTs with different sites of origin, it may be possible to identify novel prognostic biomarkers. This study aimed to identify prognostic biomarkers for GIST by comparing the molecular backgrounds and prognoses of S-GISTs and I-GISTs. A meta-analysis of 3 studies was performed to identify Differentially Expressed Genes (DEGs) between S-GISTs and I-GISTs. Receiver Operating Characteristic (ROC) analyses were performed for the DEGs, which were then ranked according to the areas under the ROC curves that they achieved. The Kaplan–Meier Method and log-rank test were used to estimate and compare survival curves. Protein-Protein Interaction (PPI) network analysis was performed using the DEGs, to identify hub genes associated with the malignant biological potential of GISTs. Overall, we identified 149 DEGs between S-GISTs and I-GISTs. Comparing I-GISTs with S-GISTs, 89 DEGs were up-regulated and 60 DEGs were down-regulated. Endothelin-3 (EDN3), which was up-regulated in I-GISTs, had the greatest area under the ROC curve of the 149 DEGs. Kaplan–Meier curves showed that high EDN3 expression was associated with significantly shorter disease-free survival in patients with S-GISTs. The EDN3 expression level was higher for high-risk patients with S-GIST than for low-risk patients with S-GISTs. Based on the PPI network of DEGs, EDN3 was identified as a hub gene related to the malignant biological potential of GIST. In summary, our study identified genes that are differentially expressed between S-GISTs and I-GISTs. EDN3 expression may have prognostic value for patients with GIST. Further validation studies of EDN3 are warranted for clinical application.

Introduction

Gastrointestinal Stromal Tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract [1]. Most GISTs show aberrant activation of Tyrosine Kinase Proteins KIT or Platelet-Derived Growth Factor Receptor-α (PDGFRα) transcribed by oncogenic mutations [2]. Imatinib, a potent Tyrosine Kinase Inhibitor (TKI) for both KIT and PDGFRα, has greatly prolonged both the survival periods and the times to metastasis after complete resection in patients with GIST [3,4]. However, because approximately 50% of operative patients with GIST are cured by surgery alone, imatinib therapy may benefit only a limited number of patients [5]. Hence, there has been a longstanding need for prognostic modalities that could optimize adjuvant imatinib therapy in cases of GIST.

The stomach and small intestine are the most common primary sites of GISTs, accounting for about 60-70% and 20-30% of cases, respectively [6]. GISTs of the Stomach (S-GISTs) and GISTs of the Small Intestine (I-GISTs) have substantially different behaviors and immunohistochemical appearances [7,8]. Several previous studies showed that risks of recurrence are clearly higher for patients with I-GISTs than for patients with S-GISTs [9-12]. It is thought that approximately 20% to 25% of S-GISTs and 40% to 50% of I-GISTs are clinically aggressive [13]. Therefore, tumor site is a factor in risk stratification schemes. Research on genetic aberrations that are specific to tumors arising at certain anatomical sites could provide clues about the molecular mechanisms of the malignant behavior of GIST, thus leading to the development of prognostic biomarkers.

Previous studies have compared S-GISTs and I-GISTs, identifying genes whose expression was unique to I-GISTs. Antonescu et al. identified Differentially Expressed Genes (DEGs) between S-GISTs and I-GISTs using 28 tumor samples, and found that 16 genes had significantly elevated expression in I-GISTs [14]. Hara et al. found that 44 genes were differently expressed between S-GISTs and I-GISTs, and demonstrated that the Slit Homolog 2 (SLIT2) gene was involved in the recurrence of GISTs and associated with a poorer prognosis [15]. Okamoto et al. revealed that aberrant DNA methylation was associated with the aggressiveness of GISTs, and those patients with methylation of RECS, paired box-3, and p16 had a significantly poorer prognosis [16]. However, each of these studies was based on a single patient cohort, and their results tended to be contradictory, probably because of the small sample sizes.
In the present study, we performed a meta-analysis using publicly available GIST gene expression data from 3 independent studies. We clarified the differences in gene expression profiles between S-GISTs and I-GISTs. We found that Endothelin-3 (EDN3) was significantly up-regulated in I-GISTs, and that a high expression level of EDN3 was associated with shorter Disease-Free Survival (DFS) in patients with GIST. We also identified EDN3 as a hub gene that was related to the malignant biological potential of GIST.

**Material and methods**

**Data collection and curation**

GIST expression studies were collected by searching the Pub Med database. The following key words were used: "GIST" and "gene expression." In addition, the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) and the Array Express database (https://www.ebi.ac.uk/arrayexpress/) were also searched to ensure that relevant studies were not missed. Data deposited between January 2006 and December 2015 was included in this study. To confirm the characteristics of the gene expression datasets of GIST for meta-analysis, sample information was investigated in the GEO and Array Express databases, as well as in the original publications. The following information was curated for each dataset: sample series, tumor type, publication, DNA microarray platform, number of cases, references, and gene-expression data. We included previous studies if they met the following inclusion criteria: 1) tumor tissues were used, 2) information on anatomical location was present, 3) the datasets contained raw data files, and 4) the number of raw data items was more than 10.

**Meta-analysis of GIST gene expression datasets**

Three published and publicly available GIST datasets were obtained from the GEO and Array Express databases: GSE8167, GSE31802, and E-MATB-373. The microarray experiments performed in this study included data for 106 GIST cases. Gene-expression data analysis was performed using R software packages from the Bioconductor Project, as follows [17,18]: For affymetrix data, all samples were extracted and analyzed using the affy package [19]. Each dataset was pre-processed using the MAS5.0 algorithm [20]. For agilent data, all samples were extracted and the statistical analyses were conducted using the limma package [21]. All DEGs were identified using the limma package (fold change >1.5, p-value <.05).

**Construction of the network and characterization of hub genes**

The Search Tool for the Retrieval of Interacting Genes (STRING), an online database resource that collects comprehensive information on predicted and experimental interactions of proteins and genes [22], was used in the present study. The interactions of gene pairs in the STRING database were displayed using a combined score. The DEGs between S-GISTs and I-GISTs were mapped into Protein-Protein Interaction (PPI) networks and a combined score of >0.5 was set as the cut-off value for significant gene pairs. The PPI network was visualized using Cytoscape software (version 3.3.0; National Institute of General Medical Sciences, Bethesda, MA, USA) [23], and the hub genes were screened as nodes with connectivity greater than 5, as reported previously [24,25].

**Statistical analysis**

We selected cutoff scores based on Receiver Operating Characteristic (ROC) curve analysis. At each mRNA expression level, the sensitivity and specificity for each outcome was plotted (thereby producing an ROC curve). To select the cutoff point, we chose the score that maximized both sensitivity and specificity. The gene expression data were dichotomized into high-level and low-level groups based on the selected cutoff values. The Kaplan–Meier method was used estimate the DFS rates in each group, which were then compared using the log-rank test. ROC curves were analyzed using R Bioconductor. The Kaplan–Meier survival analysis was conducted in Graph pad Prism 7.0 (Graph Pad Software, San Diego, CA, USA). P <0.05 was considered statistically significant.

**Results**

**Process of gene expression data selection; inclusion and exclusion criteria**: The process that was used to select the probe datasets from the GEO and Array Express databases is illustrated in Figure 1. We searched for gene expression data of GISTs that had been published during 2006–2015. We identified and manually curated published and publicly available GIST datasets. Based on the search criteria, database selection initially yielded a total of 340 studies. Original microarray data of GISTs were included in 27 studies. For our further analyses, we only used datasets that had accessible, unprocessed raw data and contained the anatomical locations of GIST samples. Twenty-four studies were excluded from the final analysis because information on anatomical location was not included in these studies. Ultimately, we identified 3 datasets that met our criteria, and analyzed these datasets in the present study. Their details are summarized in Supplementary Table 1.

![Figure 1: Flow diagram of the process by which GIST microarray datasets were selected for the meta-analysis. The publicly available, raw microarray data on GISTs in the GEO and Array Express databases were downloaded and grouped for meta-analysis.](image-url)
Identification of common DEGs between S-GISTs and I-GISTs in 3 independent datasets: In order to identify the DEGs between S-GISTs and I-GISTs, the gene expression data from 3 studies were analyzed separately using the limma package in R. The identified DEGs are shown as heat maps in Figure 2. In the GSE8167 dataset, 569 genes were significantly up-regulated and 374 genes were down-regulated in I-GISTs (Figure 2A) (fold change >1.5, p <0.05). In the GSE31802 dataset, 716 genes were significantly up-regulated and 451 genes were down-regulated in I-GISTs (Figure 2B). In the E-MATB373 dataset, 603 genes were significantly up-regulated and 697 genes were down-regulated in I-GISTs (Figure 2C).

We subsequently merged the DEG results from the 3 studies, and identified the common genes. We found that 89 genes were commonly up-regulated and 60 genes were down-regulated in I-GISTs in the 3 datasets (Figure 2D and E). A list of the common gene expression data is provided in Supplementary Table 2, and all gene expression data are shown in Supplementary Table 3. ROC curve analysis was performed for the DEGs, and the DEGs were then ranked according to the Areas Under the Curves (AUCs) that they achieved. The ROC curve analysis showed that EDN3, which was up-regulated in I-GISTs, had the highest AUC value among the 149 DEGs. The AUC for EDN3 was 0.773 (Supplementary Figure 1).

EDN3 as a potential novel prognostic marker in GISTs: Among the 3 datasets used in the meta-analysis, only 1 dataset (GSE8167) included available survival data. Thus, we investigated the association between EDN3 expression and the DFS of patients with S-GISTs using the GSE8167 dataset. The cutoff value for the prediction of survival was set at 4.85, with a sensitivity of 81.8% and a specificity of 66.7% based on the ROC curve (Supplementary Figure 1). Using this cutoff value, we divided 23 cases into 2 groups showing high (n = 14) and low (n = 9) expression. Kaplan–Meier analysis demonstrated that the DFS period of patients with high expression of EDN3 mRNA was significantly shorter than that of patients with low expression (p = 0.031) (Figure 3A). To examine the gene expression levels of EDN3 in patients belonging to different risk groups, we compared the gene expression levels of EDN3 between low- or intermediate-risk and high-risk patients with S-GISTs using the gene expression data from the 3 studies. We found that EDN3 had significantly greater expression in the high-risk patients than in the low- or intermediate-risk patients with S-GISTs (p<0.05) (Figure 3B). These results suggested that EDN3 has the potential to be a novel prognostic marker in cases of GIST.

Identification of EDN3 as a hub gene by network analysis of DEGs between S-GISTs and I-GISTs: It is well known that hub genes contribute to tumor genesis and are related to prognosis [26-28]. We analyzed gene interactions and sought to identify hub genes using the DEGs between S-GISTs and I-GISTs. Based on the STRING database, a PPI network analysis of DEGs was performed. As shown in Figure 4, the PPI network consisted of 49 nodes and 45 edges. Network analysis showed that 31 up-regulated genes and 18 down-regulated genes were involved in the network. EDN3 showed high connectivity with other genes, and was identified as a hub gene involved in the network. In addition, CD34 and Neuropeptide Y (NPY) were also identified as hub genes in the network analysis (Figure 4).
Discussion

The prognosis of patients with GIST differs according to the tumor site; I-GIST shows more aggressive behavior than S-GIST. Therefore, exploring genes that are associated with the tumor site could yield clues regarding the molecular background of malignancy in GIST. The present study is the first to have investigated the differences in gene-expression profiles between S-GISTs and I-GISTs using a meta-analytic approach. We identified EDN3 as a commonly up-regulated gene in I-GISTs from 3 independent studies. Overall, our findings suggest that analyses of EDN3 may have potential for optimizing the treatment strategy for patients with GIST.

Although several studies have compared the expression profiles of GISTs according to tumor site, these studies have had inconsistent findings, which has been a problem in GIST research. A consistent solution to this problem is to determine the overlap among the many studies using different platforms and observations. The genes with findings that are consistent across studies likely show biological relevance to aggressiveness and recurrence in GIST. On the other hand, sporadically reported genes may be false positives. Thoughtful utilization of publicly available datasets could therefore accelerate the identification of molecular signatures. The meta-analysis of datasets from several independent studies is a powerful approach and may lead to more robust, reproducible, and precise predictions [29]. In the current study, we searched for gene expression data on GIST and identified publicly available GIST datasets for the period 2006–2015. Database selection initially yielded a total 340 publications, of which 27 studies had original microarray data. Ultimately, we performed a meta-analysis of the 3 studies that had accessible raw data and contained data on the anatomical locations of the GIST samples.

EDN3 is a member of endothelin family, and is a secreted vasoactive peptide that binds to the endothelin B receptor. EDN3 is essential in the development of neural crest-derived cell lineage [30]. Mutations in EDN3 or EDNRB can lead to abnormal development of the enteric nervous system [31]. Garcia et al. demonstrated that EDN3 exhibited a tumor-angiogenic response in a melanoma mouse model [32]. In ovarian carcinoma, EDN3 has been identified in a screen for Wnt/β-catenin-induced genes [33]. Sun et al. reported that EDN3 was involved in the development of cervical cancer [34]. Liu et al. demonstrated that EDN3 played an important role in maintaining cellular and molecular properties of glioblastoma stem cells [35]. This study is the first to have demonstrated that the gene expression level of EDN3 was higher in I-GISTs than in S-GISTs, and was associated with the malignant biological potential of GISTs.

The PPI network is the basic skeleton that illustrates how proteins handle their respective functions in terms of the self-organization and homeostasis of the biological system. Aberrant function of molecular agents in the network can enhance the risks of many diseases, such as cancer [36]. Moreover, hub genes are important to study because of their centrality in PPI networks. Deletion of hub genes can lead to vast and fatal effects on the integrity of the biological network [37]. In this study, we performed a PPI network analysis using the DEGs between S-GISTs and I-GISTs. Based on the PPI network, we found 3 hub genes associated with malignant biological potential: EDN3, NPY, and CD34. NPY and CD34 have been also reported in various types of cancers. Hong et al. demonstrated that high NPY release is associated with Ewing sarcoma bone dissemination [38]. Medeiros et al. reported that NPY stimulated proliferation and migration in breast cancer [39]. CD34 was reported to be an immunohistochemical marker for differential diagnosis in cases of GIST. CD34 expression is positive in about 80% of S-GISTs, but only in about 50% of I-GISTs [40]. In our study, we found that CD34 had significantly greater expression in S-GISTs than in I-GISTs, and these results are consistent with previous research.

Several limitations should be considered in this study. First, the results of meta-analyses are dependent on the reliability of the original data, which was not validated in this study. Second, we did not examine the molecular functions of EDN3 in vitro or in vivo. Currently, no GIST cell line is available in public cell banks. Further analysis of the functional significance of EDN3 will lead to a more detailed understanding of the disease mechanisms of GIST, potentially helping to reveal novel therapeutic modalities.

In conclusion, we demonstrated differential expressions of 149 genes associated with tumor site in this meta-analysis of cases of GIST. We also found that 1 of these DEGs, EDN3, was significantly unregulated in I-GISTs. Prognostication using EDN3 may help to optimize treatment strategies for patients with GIST.

Acknowledgements

This research was supported by a Practical Research for Innovative Cancer Control grant (15ck0106089h0003) from the Japan Agency for Medical Research and Development, AMED.

References


