Original article

Gene co-expression analysis and Network biology studies in Indian population reveals functional similarities between Gastric cancer and other metabolic disorders

Blessantoli Mohandas¹, J. Jannet Vennila², Nikhil Ruban³

Abstract:

Objective: Gastric cancer (GC) is a multifactorial disease and known to have been associated with metabolic disorders. Gene expression profiling among various GC populations will help to strategize diagnosis and treatment. The current study employed microarray data analysis (MDA) and network biology methods to understand the significant genes in a GC Indian population and its association with other metabolic disorders. Materials and Methods: The microarray datasets of GC Indian population (Bangalore) was retrieved from Gene Expression Omnibus (GEO), normalized and analyzed using GeneSpring. With the fold change of ± 1.5, the differentially expressed genes (DEG) were identified. An interactome was built to study interactions and generate gene clusters. Statistical (centrality) parameters were applied to identify highly connected clusters followed by functional enrichment to identify significant pathways associated with the GC genes. Results and Discussion: MDA identified 7181 DEGs (3984 up regulated genes and 3197 down regulated) and the interactome yielded 16552 interactions and two sub clusters. Cluster 1 was found to be statistically fit. The functional characteristics of the significant genes in this cluster revealed their association with adrenal cortex hormone insufficiency, thyroid disorders and deficiencies in kidney water resorption. Conclusion: It is inferred from our study that, deficiency in Thyroid, Adrenal hormones and Antidiuretic hormone (ADH) functions has fair share in the prognosis and pathogenesis of GC Indian population. Henceforth, GC should not be viewed as separate entity in the series of cancers and gene expression profiling will help in improvising personalized medicine.

Keywords: Gastric Cancer; Network biology; Microarray Data Analysis; Metabolic disorders and Gastric Cancer; Gene expression profiling

Introduction

Gastric cancer (GC) or Stomach cancer is the fifth most predominant cancer in the world, accounting with 7,23,000 cancer death rate per year and the third most prevalent cause of cancer-related mortality¹. Approximately 50% of GC people live in East Asia, Southern Asia and China, primarily may be due to poor dietary patterns and Helicobacter pylori infection². In the early phase, GC is asymptomatic and therefore 80 to 90% of GC patients are diagnosed only in advanced stage. Defining new molecular genetic markers through gene expression profiling would help in better understanding of the GC prognosis. The GC gene expression pattern varies among populations depending on genetic makeup, nutritional status and exposure to environmental factors.

1. Blessantoli Mohandas, Department of Biochemistry, School of Agricultural and Biosciences, Karunya Institute of Technology & Sciences, Coimbatore-614114.
3. Nikhil Ruban, Department of Humanities and Social sciences, Birla Institute of Technology and Sciences, Pilani Campus, Rajasthan-333 031.

Correspondence: Dr. J. Jannet Vennila, Director, Shift II and Coordinator, MCC MRF Innovation Park, Madras Christian College, Chennai-600059. Email- jannetvennila@gmail.com
dietary pattern and environmental influence. The genes involved in other metabolic pathways can also be deemed to be significant in GC pathogenesis and progression. Although records have shown 57,394 increasing new cases in India, there are only very few data available on gene expression profiling in GC Indian population. This study focused on systematic evaluation of expression and network analysis of significant genes in GC and maps their relevance to important molecular/metabolic mechanism. This will help to strategize diagnosis, development and treatment process in GC Indian population.

**Method**

**Data selection:** Microarray dataset of GC, Bangalore, India was retrieved (GSE 22804) from Gene Expression Omnibus (GEO). The dataset comprised of surgically resected GC samples and their paired disease-free peripheral nonmalignant tissue. Figure 1 gives the flowchart of the methodology adopted out in this study.

**Microarray data analysis (MDA):** The Microarray Data Analysis (MDA) was carried out using GeneSpring GX 12.6.1, an Agilent microarray data analysis software. Normalization of the dataset was performed using Percentile shift algorithm for Agilent platform array. The differentially expressed genes (DEGs) were identified (Fold change of ±1.5) and were categorized into ‘up’ and ‘down’ regulated based on fold change.

**Network studies:** The gene-gene correlation of all the DEGs was studied using Cladist. The Interactome of the DEGs was built using Cytoscape 3.2.0 (http://cytoscape.org). With the help of GC marker genes reported in earlier studies (TP53, IL1B, IL1RN, VEGFA, TNF, CDH1, CLDN1, CLRN3, INHBA, and SULF1), their first neighboring genes were selected in the network and a subnetwork was built. The interactions were further enriched using Bisogenet (Cytoscape plugin). Molecular Complex Detection (MCODE) was implemented to identify densely connected regions in GC networks (subclusters). K-core algorithm was used to identify the highly interconnected nodes (genes) in the network using MCODE. Centrality measures were used to predict the hub genes (highly connected genes) and to identify the node’s functional gene. Using the Cytoscape plugin Network Analyzer, the topological and statistical understandings of each sub-clusters were calculated. Centrality parameters - betweenness centrality (BC), closeness centrality (CC), Cluster Coefficient (CLC), degree and topological centrality (TC) were considered to validate the prominence of biological network to predict highly connected genes which are denoted as “significant genes”. WebGestalt and ToppGene were used to examine the functional characteristics of the significant genes. The pathways which represent many of the significant genes were studied to establish their correlation with other metabolic functions.

**Ethical clearance:** Not applicable

**Results**

**Microarray Data Analysis:** GC dataset of Indian Population (GSE22804) contained 28 samples (14 GC patients and 14 normal of the same patient). Grouping of the dataset was done as ‘normal’ and ‘diseased’. Based on the fold change cut off (±1.5) and p-value (0.05) the differentially expressed genes (DEGs) were selected. 3984 up regulated genes and 3197 down regulated genes were identified.

**Network construction and analysis:** The above
DEGs were analyzed for gene-gene correlation which revealed additional 15,586,704 co-expressed genes for up regulated and 10,220,809 for down regulated genes respectively. Among these genes, 5611 upregulated and 1,92,156 downregulated genes were observed to be statistically correlated (r value ≥ 0.9). These genes were merged together to build the interactome. The network was extended with enrichment of additional interactions using Bisogenet. And the network enriched with 16552 edges (interactions).

**Identification of highly connected sub-clusters in interactome:** K-core algorithm (MCODE) was used to identify the highly interconnected nodes (genes) in the network and two highly interconnected clusters were extracted from the interactome.

**Topological and Centrality measures:** Based on the topological and centrality measures, the cluster 1 was found to be statistically fit with R^2 value ≥ 0.9 and predicted to possess more number of functional and significantly relevant genes.

**Functional enrichment for the significant cluster:** The functional enrichment of the cluster 1 genes was performed further for identifying the significant pathways associated with the GC genes (Table 1).

**Table 1: Significant pathways associated with GC genes**

<table>
<thead>
<tr>
<th>Pathways</th>
<th>Total genes</th>
<th>Genes common to GC</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid hormone synthesis</td>
<td>72</td>
<td>5</td>
<td>0.0016</td>
</tr>
<tr>
<td>Autoimmune thyroid disease</td>
<td>47</td>
<td>3</td>
<td>0.0018</td>
</tr>
<tr>
<td>Proximal tubule bicarbonate reclamation</td>
<td>23</td>
<td>3</td>
<td>0.0025</td>
</tr>
<tr>
<td>Toll-like receptor signaling pathway</td>
<td>99</td>
<td>4</td>
<td>0.0031</td>
</tr>
<tr>
<td>Adrenal cortex hormone insufficiency pathway</td>
<td>147</td>
<td>5</td>
<td>0.0031</td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td>12</td>
<td>2</td>
<td>0.0087</td>
</tr>
<tr>
<td>Bile secretion</td>
<td>68</td>
<td>4</td>
<td>0.0088</td>
</tr>
<tr>
<td>Vasopressin-regulated water reabsorption</td>
<td>42</td>
<td>3</td>
<td>0.0137</td>
</tr>
<tr>
<td>Endocrine and other factor-regulated calcium reabsorption</td>
<td>46</td>
<td>3</td>
<td>0.0175</td>
</tr>
<tr>
<td>Cytosolic DNA-sensing pathway</td>
<td>57</td>
<td>3</td>
<td>0.0307</td>
</tr>
</tbody>
</table>

a Adrenal cortex hormone insufficiency disorders; b thyroid disorders; c deficiency in kidney water reabsorption. Statistical significance of P<0.05

It can be inferred from Table 1, that genes associated with GC are also significant in pathways associated with Adrenal cortex hormone insufficiency (5 genes), thyroid disorders (8 genes) and deficiencies in kidney water reabsorption (6 genes). The possible role of these GC genes in interconnecting with the above metabolic disorders is diagrammatically explained in Figure 2.

APC-Adenomatous polyposis coli, SOD-Superoxide dismutase, ADH-Antidiuretic hormone, ATP1B4-ATPase Na+/K+ Transporting Family Member Beta 4, TSHR - Thyroid Stimulating Hormone Receptor, IFNA8- Interferon Alpha 8, SLC26A4- Solute Carrier Family 26 Member 4, GNAS- Guanine Nucleotide Binding Protein (G Protein) Alpha Stimulating Activity Polypeptide, PDE11A- Phosphodiesterase 11A, KMT2A- Lysine Methyltransferase 2A, WNT8B- Wnt Family Member 8B, AIPL1- Aryl Hydrocarbon Receptor Interacting Protein Like 1, Dync2h1-Dynein Cytoplasmic 2 Heavy Chain 1. Red arrows represent up-regulation and Green arrows represent down-regulation

**Discussion**

**Gastric cancer and metabolic disorders: Functional Similarities**

**Gastric cancer and thyroid disorders:** The risk factors in GC are associated with thyroid disease and vice versa. A study by Carvalho and Fighera has shown that the prevalence of gastrointestinal diseases increased in patients with autoimmune thyroid dysfunction. Similarly, in autoimmune thyroid disorder celiac disease the atrophic gastritis are highly prevalent. It is assumed that the development of certain gastric cancers involves dietary iodine which is also related to thyroid dysfunction. The reason could be morphological and functional similarities of stomach and thyroid tissue which use the membrane active transport movement mechanism to concentrate iodides. And thyroid hormone is also a potent cofactor for tumor-suppressing genes.

**Gastric cancer and deficiency in Kidney water reabsorption:** Inadequate antidiuretic hormone (ADH) leads to hyponatremia and excessive water reabsorption. Prolonged nausea and vomiting in hyponatremia may disturb the digestive system and induce GC.

**Gastric cancer and Adrenal cortex hormone insufficiency:** Chabre and his research team have observed that the abnormal expression of gastric inhibitory polypeptide (GIP) receptor enables...
adrenocortical cells to react to food intake, with a cAMP increase in which both cortisol secretion and tumor proliferation can be stimulated\textsuperscript{43}. GIP and Glucagon can regulate the gastric acid secretion in humans. GIP is also a potent releaser of gastric somatostatin, a secretion acid inhibitor \textit{in vivo} and \textit{in vitro}\textsuperscript{44}. GIP receptor may also play a role in the development of tumor cells. The GIP effects were similar to the effects of ACTH on tumor cells.

**Conclusions**

Gastric cancer (GC) is a heterogeneous disease known to associate with environmental and genetic predisposing factors\textsuperscript{45}. And from our network biology study in Indian population and reports on previous literatures, it was evident that, GC has close association with other metabolic disorders. Deficiency in Thyroid, Adrenal hormones and Antidiuretic hormone (ADH) functions has been found to have distinguished share in the prognosis and pathogenesis of GC. It is suggested that GC should not be viewed as separate entity in the series of cancers. GC may arise due to defective functions of the above said metabolic disorders. Therefore, a, multilevel screening of metabolic functions and gene expression profiling is suggested before the treatment of GC, could pave a better solution in treatment strategy. In addition, as genetic predisposition is an important feature in GC, the markers may not be universal for all the patients worldwide. Populations wise gene expression profiling can be more appropriate. It is recommended that GC gene expression profiling of different populations within India could still validate the above findings.

**Conflict of interest statement:** The authors declare that they have no conflict of interests.

**Funding:** Nil

**Authors’ contributions**

Blessantoli Mohandas acquired data and analysed the data. J.Jannet Vennila designed the study and drafted the manuscript for important intellectual content. Nikhil Ruban assisted in manuscript preparation, editing and review. The manuscript was read and approved by all the authors.

**Acknowledgments**

The authors wish to thank the technical support received from the Department of Biotechnology, Karunya Institute of Technology & Sciences, Coimbatore, India.
References:


27. Kumar S, Kumar A and Dixit VK. Evidences showing


