Abstract: Gastric cancer (GC) is one of the most lethal, as well as one of the heterogeneous, cancer types. Possible GC molecular mechanisms could be revealed by mutational co-occurrence analyses. Despite a known association between mutational co-occurrences and GC signalling contexts, no specific mechanisms have been identified. Here, known GC signalling contexts, including cancer hallmarks (DNA repair, WNT signalling, Notch signalling), were inspected in terms of mutational co-occurrences, and in particular, for a specific GC phenotype, microsatellite status (stable or low or high instability). By correlating mutational co-occurrences of gene pairs within cancer hallmarks, we constructed mutational co-occurring networks for each type of microsatellite status. As a result, we found that one status type, microsatellite-stable (MSS), associated with mutation of \textit{JAG1}, likely co-occurs for genes belonging to the WNT and Notch signalling pathways. Our study may support the feasibility of a new therapeutic strategy of designing compounds that target Notch signalling, in MSS GC patients.

Keywords: gastric cancer; microsatellite stability status; network models; association index; Notch; Wnt; DNA repair.


Biographical notes: Sungjin Park received his PhD degree in Biomedical Engineering from Inje University in 2016. Currently, he is a postdoctoral researcher in Gachon Institute of Genome Medicine and Science at Gachon Univ. Gil Medical Center. His research interests include formalisation of genomic information, computing effect of mutational interaction, cellular compartments interaction, and computation of genotype, ribotype, and phenotype interactions in cancer systems biology.

Seungyoon Nam received his PhD degree in Bioinformatics from Seoul National University in 2008. From 2011 to 2015, he worked as a senior researcher in New Experimental Therapeutics (NExT) Branch in National Cancer Center, South Korea. Currently, he is an assistant professor at the College of Medicine of Gachon University, South Korea. His research area includes systems biology, noncoding RNAs, and genomic therapeutic targeting for various diseases.
1 Introduction

Gastric cancer (GC) is one of the highly prevalent cancer types in East Asia including Korea, Japan, and China (Cancer_Genome_Atlas_Research_Network, 2014; Cristescu et al., 2015). Also, a recent, large genomics study revealed a wide, strongly heterogeneous mutational spectrum in GC (Cortes-Ciriano et al., 2017).

Recent GC studies by our group (Chang et al., 2016a; Nam et al., 2014) indicated that \textit{WNT5A} significantly associated with GC prognosis; we also demonstrated the feasibility of therapeutically targeting WNT signalling (Chang et al., 2016b). In addition, Notch signalling, a well-known cancer hallmark (Hanahan and Weinberg, 2011), also represents a new, promising therapeutic target in GC (Yao et al., 2017), and DNA mutations have been intimately associated with DNA repair pathways (Hanahan and Weinberg, 2011).

Although recent studies (Cui, 2010; Deng et al., 2012) suggest mutational co-occurrences in the pathogenesis of GC, the above-mentioned GC signalling contexts have yet to be considered in such analyses.

In this study, we inspected specific mutational co-occurrences and signalling contexts in GC, by constructing mutational correlation networks which provide systematic views of complex systems (Arredondo et al., 2012; Kim et al., 2014; Mallavarapu et al., 2016; Park and Nam, 2017) and allow to comprehend interactions among cellular components (Fayruzov et al., 2011). To that end, we applied an association index to transform a bipartite matrix (Fionda et al., 2009) (phenotype status by mutation status matrix) to a monopartite matrix (correlation of mutational co-occurrences between two genes) in the context of the three cancer hallmark signal pathways: DNA repair, WNT signalling, and Notch signalling. This approach allowed us to identify co-occurring mutations, and their influence, by measuring a degree of interconnection.

We then analysed our mutational co-occurrence network using The Cancer Genome Atlas (TCGA) (Tomczak et al., 2015) stomach adenocarcinoma (STAD) dataset, and chose the disease phenotype of microsatellite status, defined in the TCGA by the clinicopathological features: MSI-H, as microsatellite instability-high; MSI-L, microsatellite instability-low; and MSS, microsatellite stable.

2 Methods

Our mutational co-occurrence network was built using three steps; (1) extracting subsets of a bipartite matrix; (2) comparing mutational co-occurrence profiles; and (3) constructing networks (Figure 1).
Figure 1  Framework for constructing a mutational co-occurrence network

We first used the TCGA STAD somatic mutation dataset (version: TCGA_STAD_mutation_curated_broad_gene-2015-02-24; curated by Broad Institute Genome Sequencing Center) from the UCSC Cancer Genomics Browser (CGB) (Zhu et al., 2009) website. From a total of 289 patients in the STAD somatic mutation dataset, the number classified as MSI-H was 63, the number classified as MSI-L was 44, and the number classified as MSS was 81.

We then selected three cancer hallmark signal pathway gene sets; the DNA repair pathway, the Wnt/beta-catenin signalling pathway, and the Notch signalling pathway. Considering that MSI is believed to result from impaired DNA mismatch repair (Cortes-Ciriano et al., 2017), we first explored mutational co-occurrences of genes in this pathway. MSI-H is known to be less metastatic than MSS in cancer (Gryfe et al., 2000), and we previously demonstrated that the Wnt/beta-catenin signalling pathway is important for STAD tumourigenesis, via computational analysis, in vitro assays, and xenograft models (Chang et al., 2016a; Nam et al., 2014; Park et al., 2016). Moreover, both the Wnt/beta-catenin and Notch signalling pathways associate with cancer stem cells (CSCs) (Lamouille et al., 2014), and in a xenograft model, inhibiting Notch not only reduced CSC numbers but also suppressed the epithelial-mesenchymal transition (EMT) (Smith et al., 2012). We extracted subsets of genes from the DNA repair, Wnt/beta-catenin, and Notch signalling pathways, from the STAD mutation dataset, using hallmark collections from the MSigDB database (Liberzon et al., 2015), and the terms `HALLMARK_DNA_REPAIR`, `HALLMARK_WNT_BETA_CATENIN_SIGNALING`, and `HALLMARK_NOTCH_SIGNALING`.

Next, we performed association indexing using Pearson correlation coefficients (PCCs), for each gene pair, to calculate the correlation between the mutational co-occurrence profiles of the genes A and B, as:

\[ PCC_{AB} = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B}, \]

where \( \text{cov}(A, B) \) is the covariance between the expression levels of genes A and B, and \( \sigma_A \) and \( \sigma_B \) are the standard deviations of the expression levels of genes A and B, respectively.
where \( \text{cov} \) is the covariance, and \( \sigma \) is the standard deviation. From this, we built a monopartite matrix of STAD mutational co-occurrences, setting criteria according to Green et al., (Green et al., 2011); PCC > 0.7, with \( p < 0.05 \); and PCC > 0.46, with \( p < 0.05 \). PCCs and their respective p-values were calculated using the function ‘pearsonr’ in SciPy’s stat module (Jones et al., 2001). We then constructed networks of monopartite matrices using Cytoscape version 3.4 (Shannon et al., 2003).

3 Results and discussion

Overall, we profiled mutational frequencies belonging to the DNA repair, Wnt/beta-catenin signalling, and Notch signalling pathways, according to MS status. The number of patients per group was 63, 44, and 181, for MSI-H, MSI-L, and MSS, respectively. As expected, MSI-H patients exhibited more hypermutation than MSI-L and MSS patients. The average number of mutations per patient was 6.94 (MSI-H), 1.2 (MSI-L), and 1.02 (MSS), for genes belonging to the DNA repair pathway. For Wnt/beta-catenin signalling pathway component genes, the averages were 5.03 (MSI-H), 2.09 (MSI-L), and 1.8 (MSS); for the Notch signalling pathway, however, average mutation numbers dropped to 2.49 (MSI-H), 0.3 (MSI-L), and 0.19 (MSS), respectively (Table 1).

Table 1 Average numbers of mutated genes per GC pathway

<table>
<thead>
<tr>
<th>Pathway</th>
<th>MSI-H</th>
<th>MSI-L</th>
<th>MSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA repair</td>
<td>6.94</td>
<td>1.2</td>
<td>1.02</td>
</tr>
<tr>
<td>Wnt/beta-catenin signalling</td>
<td>5.03</td>
<td>2.09</td>
<td>1.8</td>
</tr>
<tr>
<td>Notch signalling</td>
<td>2.49</td>
<td>0.3</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The DNA repair pathway, with regard to MSI-H status, was more mutation-loaded than previously reported (Cortes-Ciriano et al., 2017). For MSI-L and MSS statuses, the mutational load was higher for the Wnt/beta-catenin signalling pathway, as compared to the other two pathways of interest. Moreover, in contrast to MSI-H, mutational occurrence rates were barely observed in MSI-L and MSS for the Notch signalling pathway. Figure 2 shows our examination of mutational occurrence ratios for the three hallmark pathways, according to MS status, showing the top 15 genes for each pathway. For example, TP53 has been reported that an average rate of mutation incidence is about 50% in gastric cancer. In MSI-H, however, the ratio was about 10% lower than average, while it was 10% higher for MSI-L patients.

We next applied the above-described procedures to the TCGA STAD mutation matrix, to generate monopartite matrices that link mutations whose profiles correlated. For MSI-L, using our association index’s criteria (PCC \( \geq 0.7 \) and p-value \( < 0.05 \)), only the DNA repair monopartite matrix was derived, while both the DNA repair and Wnt/beta-catenin monopartite matrices were produced for MSI-H. The DNA repair monopartite matrix for MSI-H identified eleven modules, with forty-seven nodes and sixty-seven edges, while the Wnt/beta-catenin monopartite matrix for MSI-H identified only one module with two nodes and one edge. The DNA repair monopartite matrix for MSI-L showed two modules, with five nodes and four edges. For MSS, the monopartite...
matrices of DNA repair, Notch signalling, and Wnt/beta-catenin signalling showed a DNA repair monopartite matrix having six modules, with twenty-five nodes and seventy-nine edges, the Notch signalling pathway having two modules, with eight nodes and thirteen edges, and the Wnt/beta-catenin signalling matrix having four modules, with ten nodes and nine edges.

**Figure 2** The top 15 most prevalent gene mutations in three signal pathways (DNA repair, Wnt/beta-catenin, and Notch). Ordering of the gene lists was by summation of total mutational occurrences of all three microsatellite statuses.

We then built MSS, MSI-H, and MSI-L, unidirectional networks (Figure 3) from their respective six one-mode matrices. Based on these, we measured network parameters (network density, network heterogeneity, network centralisation, and average number of neighbours) for the three networks of interest (Table 2).

<table>
<thead>
<tr>
<th>Network</th>
<th>Density</th>
<th>Heterogeneity</th>
<th>Centralisation</th>
<th>Average number of neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSS</td>
<td>0.117</td>
<td>0.942</td>
<td>0.184</td>
<td>4.81</td>
</tr>
<tr>
<td>MSI-H</td>
<td>0.057</td>
<td>0.854</td>
<td>0.136</td>
<td>2.735</td>
</tr>
<tr>
<td>MSI-L</td>
<td>0.4</td>
<td>0.306</td>
<td>1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

For the MSS network (Figure 3), **JAG1** represented a crosstalk node between the Wnt/beta-catenin and Notch signalling nodes. Its co-occurring partner nodes were **CSNK1E**, **RB1**, and **SKP2**, in the Wnt/beta-catenin signalling pathway, and **PSENEN**, **SAP30**, and **WNT2**, in the Notch signalling pathway. Moreover, **JAG1** was previously reported as linked to the Wnt/beta-catenin and Notch signalling pathways pathologically (Rodilla et al., 2009), to induce metastasis in lung cancer (Chang et al., 2016c), and is considered a promising drug target for antitumor therapy (Li et al., 2014). Thus, our crosslinking of **JAG1** in a mutational co-occurring network might be key to understanding why metastasis is more frequent in MSS than MSI-H.
Figure 3  Mutational co-occurring networks for the three phenotypes (MSS, MSI-H, and MSI-L). Node sizes were determined by network property values using between-centralities. In the MSS network, JAG1 represented a crosstalk node between the WNT/beta-catenin and Notch signalling networks. Networks shown in the upper panel were for Pearson correlation coefficients (PCCs) > 0.7, while networks in the lower panel were for Pearson correlation coefficients (PCCs) > 0.46.

While MSI-H was the most hyper mutated (an average mutational occurrence per patient of 6.94 for the DNA repair pathway, Table 1), the density of the MSI-H network was only 0.057 (Table 2). Comparing this to the MSS network density (0.117, and a mutational occurrence average of 1.02 for the DNA repair pathway), the mutational co-occurrence degree of MSI-H might be less than MSS, even despite the higher mutation ratio. Thus, this result might be interpreted such that MSI-H tends to cause mutagenesis more randomly than MSS, but this hypothesis requires more inspection. Via our construction of mutational co-occurrence networks, using different PCC thresholds, and as presented in Figure 3, the relevant connections varied between different neighbour nodes. For example, with a PCC threshold of 0.7, Notch and Wnt/beta-catenin signalling pathway were connected by JAG1, as mentioned above; however, DLL1 was also a crosstalk node between those two pathways, with a PCC threshold of 0.46. The lower threshold correlation coefficient could allow identification of more functional connections, even while the variance was high. Strictly, high threshold correlation coefficients might allow for more reliable pairs of nodes, although it is possible to yield an empty result missing functionally meaningful connections. Therefore, one must measure the significance of a network within a specific context. To that end, the Connection Specificity Index (CSI) (Green et al., 2011) might help overcome these limitations.
Additionally, we downloaded the oesophagus-stomach cancer dataset (Cancer Genome Atlas Research et al., 2017) with 559 samples from cBio Portal that published in 2017. We chose only missense mutations, then calculated Pearson correlation coefficient values of genes in Wnt/beta-catenin signalling pathway and Notch signalling pathway, and presented resultant networks with PCC cut-offs 0.7 and 0.46 (Figure 4; omitting orphan nodes here). With PCC > 0.46 criterion, JAG1 and DLL1, which were crosstalk nodes between Wnt/beta-catenin and Notch signalling pathways, were identified in similar manner of our result in Figure 3. However, with PCC threshold 0.7, PSEN, SAP30, PRKCA, and WNT2 were appeared while JAG1 were disappeared. Because PCC values of JAG1-CSNK1E, JAG1-PSENEN, and JAG1-SAP30 were 0.575 each.

Figure 4  Missense mutation co-occurring networks for the oesophagus-stomach cancer MSS type samples. In the MSS network, JAG1 and DLL1 represented a crosstalk node between the WNT/beta-catenin and Notch signalling networks with PCC threshold at 0.46, but both genes were disappeared in PCC threshold 0.7. Networks shown in the left panel were for Pearson correlation coefficients (PCCs) > 0.7, while networks in the right panel were for Pearson correlation coefficients (PCCs) > 0.46. Node sizes were determined by network property values using between-centralities.

4 Conclusion

In this work, we generated a network model that can effectively evaluate an association between phenotypes and mutated genes in cancer. Also, compared to traditional clustering, this network method is powerful enough to cluster co-occurring mutations, within specific signal pathway contexts.
Acknowledgements

This research was also supported by the Korea Institute of Science and Technology Information (KISTI) (to S.N.), and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03032077 to S.P.; NRF-2016R1D1A1B03933145 to S.N.).

References


A mutational co-occurrence network in GC based on an association index


