GSJ: Volume 9, Issue 8, August 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

Analysis of symptom clusters cross different diseases based on Biclustering

Edouard Niyongabo [a], Patrick Habimana [b]

a *Department Information technology-Networking, Faculty of computing and information sciences, University of Lay Adventists of Kigali, Rwanda*

b *Department Information technology-Networking, Faculty of computing and information sciences, University of Lay Adventists of Kigali, Rwanda*

Corresponding Email: [a] niyongabo2020@gmail.com, [b] habimanapat7@gmail.com

Abstract. The symptoms are important signs of disease diagnosis and treatment. Medical researchers have found that some Symptom often occur simultaneously in different diseases and patients, forming Symptom Cluster, a Symptom Cluster with significant co-occurrence pattern. The traditional symptom group analysis method mainly consists of the principal component and hidden class model. However, in the symptom group analysis of cross-disease species, the association of disease information is often needed, therefore, there are significant deficiencies. Biclustering method was used to analyze the symptoms according to their manifestations associated with various diseases, grouping of diseases based on the associations of different symptoms. We extend previous research by describing and implementing algorithms to identify subgroups of diseases and subgroups of symptoms co-currently, by performing simultaneous clustering of both rows and columns in data matrix. To determine the underlying shared molecular mechanisms (in terms of shared genes and shortest paths of protein interactions) of symptom cluster using symptom gene relationship data. Using the data of disease symptom relationship with whole disease spectrum, the structure and molecular mechanism of specific symptom group was identified. In this study, we used data from 16383 disease symptom relationships (including 13532 diseases and 2378 symptoms). Three classical Biclustering algorithms, BIMAX, QUBIC and Spectral Clustering, are used to cluster the symptom groups.

Further molecular correlation analysis using the interaction data from 371422 symptom gene relationships and 841068 records (from STRING 11, a large-scale integrated database) These symptom groups share the molecular mechanism such as shared gene and short interaction group molecular pathway ($p < 0.05$) compared with random symptom combinations.

Among the three methods, we found that the QUBIC method formed the clustering result with good biological significance, while the spectral clustering obtained the best symptom clustering result in sharing the molecular mechanism. In the follow-up study, the performance of different methods for symptom group analysis can be further explored.

Key words: symptom cluster, Biclustering, symptom, gene association, protein-protein interaction network.

1. INTRODUCTION

The identification of symptom clusters in medicine and illness care is not new. In fact, from the Middle Ages to the late 19th century, symptoms were generally thought of as the bodily or mental phenomena that constituted specific illnesses. In the 20th century, it became known and accepted that underlying pathophysiologic mechanisms were responsible for the pattern of symptoms that typified different diseases. As the diagnosis of disease became more sophisticated, symptom-based diagnostic criteria were supplanted by laboratory and imaging tests and symptom clusters received less attention[1].

Patients with chronic conditions, such as cancer and other rare diseases, experience an array of multiple co-occurring symptoms (e.g., pain, fatigue, sleep disturbance). When these symptoms remain under diagnosed and undertreated, they have a negative impact on patient-reported outcomes (PROs) including functional performance, cognitive status and quality of life [2],[3].

A reduction in symptom burden in these patients has the potential to improve their capacity to live well over their entire lives [4]. To achieve this goal, a transformation is needed in how multiple co-occurring symptoms are assessed and managed in order to improve patient outcomes and stimulate a reduction in health care utilization and costs[5].

A strategic plan that advances symptom science through symptom cluster research has the potential to accelerate the growth of an empiric body of knowledge that is capable of sustaining innovative symptom management interventions in these patients [6]. While research often focuses on a single symptom, in cancer and most other chronic conditions, patients experience multiple co-occurring symptoms that are related to each other (i.e., symptom clusters). Compared with a single symptom, the occurrence of symptom clusters appears to worsen patient outcomes [6], [9].

The science of symptom clusters and its application to practice should be important to clinicians for three central reasons. First, evidence indicates that symptom clusters warn of negative outcomes such as depression, functional or role limitations, poorer quality of life and mortality. Ignoring symptom clusters may jeopardize important patient health outcomes. Second, knowledge of symptom clusters allows for more thorough symptom assessment. If clinicians are aware of symptoms that typically co-occur, then when a problematic symptom is identified through standard symptom assessment procedures, clinicians can anticipate and probe further into other likely related symptoms. This may result in more efficient use of limited patient-provider time and potentially uncover symptoms that might otherwise have been overlooked. Third, recognizing the co-occurrence of specific symptoms creates the possibility of more efficient symptom management by targeting the cluster of symptoms with a single treatment approach[10].

Our work is motivated by unrelieved symptoms can have deleterious effects on patient outcomes. Patients with chronic diseases experience a variety of symptoms as a result of their disease or as a result of treatments for their disease [11][13]. These symptoms are a major problem for patients, as well for their family caregivers, because the management of these symptoms is often the responsibility of the patients themselves [5].

Consequently, gaps in knowledge exist regarding the clinical meaning of symptom clusters, the specific symptoms which may cluster, and the reasons for clustering. Even if a continuing focus on single-disease research is crucial, it is equally important that symptom management research begin to evaluate multiple symptoms cross different disease conditions because treating one disease may not necessarily improve quality of life.

2. RELATED WORK

Analysis on symptom clusters in not new because it is shown in below table. Totally different strategies are exploitation to cluster symptoms, by reading previous works done on symptom clusters can show the gaps in current analysis. The prevalence of symptom clusters studies investigated about different types of cancer[19]–[21][10].

Author, Year, Title, Purpose and Design	**Symptom assessment instrument(s), number of symptoms on instrument; statistical analysis method, symptom dimension(s) used to create symptom clusters**	**Number of symptom clusters, specific symptoms within each cluster**	**Strengths and limitations**
Chen et al., 2007[22] ***Title:*** cancer symptoms clusters: A validation study ***Purpose(s):*** To validate the 3 factor symptom structure by using CFA in a larger sample of cancer patients and to examine how 4 disease/treatment variables (diagnosis, disease stage, cancer treatment, hospitalization) and one outcome variable (functional status) were associated with the 3 symptom factors (sickness symptoms, GI symptoms, emotional symptoms) ***Design:*** Cross-sectional	***Instrument(s):*** MDASI-T: 13 items ***Analysis:*** CFA with maximum likelihood estimation. 9 MDASI symptoms used to build the measurement model ***Dimension(s):*** Severity	3 symptom clusters identified: Sickness symptoms: pain, fatigue, disturbed sleep, lack of appetite, drowsiness GI symptoms: nausea, vomiting Emotional symptoms: distress, sadness	***Strengths:*** Evaluated relationships between symptom cluster "scores" and disease and treatment characteristics. Demonstrated that higher symptom cluster "scores" were associated with decreased functional status ***Limitations:*** Used only 9 symptoms from the MDASI-T Heterogeneous cancer diagnoses Cross-sectional design
Karabulu et al., 2010[12]	**Instrument(s):**	3 symptom clusters identified:	***Strengths:***

Title: Symptom Clusters and Experiences of Patients with Cancer ***Purpose(s):*** To characterize the prevalence and severity of symptoms in Turkish patients with cancer and describe the clustering symptoms ***Design:*** Cross-sectional	MDASI: 13 items ***Analysis:*** Hierarchical cluster analysis ***Dimension(s):*** Severity Interference	Cluster 1: general activity, mood, work, relations with other people, walking, enjoyment of life. Cluster 2: sleep disturbance, difficulty in remembering, pain, distress, sadness, fatigue, dry mouth, appetite loss Cluster 3: nausea, vomiting, shortness of breath, numbness, drowsiness	First study in Turkish oncology patients Relatively large sample size ***Limitations:*** Cross-sectional design Inclusion of interference items in the symptom cluster analysis Symptom clusters were not named MDASI assesses only 13 items
Gift et al., 2007[23] ***Title:*** Symptom clusters in patients with lung cancer: A literature review ***Purpose(s):*** To determine whether symptoms co-occur in patients newly diagnosed with lung cancer; whether symptoms vary according to antecedents of stage of disease, comorbidities, treatment, or gender and; whether co-occurring symptoms affect performance ***Design:*** Cross-sectional	**Instrument(s):** Medical Outcomes Study (SF-36): only 16 items used in this analysis Physical Symptom Experience: 37 items ***Analysis:*** Exploratory maximum likelihood factor analysis ***Dimension(s):*** Occurrence Severity	Only 1 stable symptom cluster was identified: Un-named cluster with 7 symptoms: nausea, fatigue, weakness, appetite loss, weight loss, altered taste, vomiting	***Strengths:*** Evaluated a homogeneous cancer type with a varying stage of disease Evaluated symptoms in elderly patients ≥ 65 Used physical symptom scales that are known to primarily address the physical dimension of symptoms ***Limitations:*** Did not find stable clusters in 3 of the 4 factor loadings Analysis consisted of predominantly white males The effects of age on functional status were not controlled for in this study Symptoms cluster groupings were different from other studies

Molassiotis et al., 2010[24] ***Title:*** Symptom Cluster Patterns During the First Year After Diagnosis with Cancer ***Purpose(s):*** To explore the patterns of clusters over time, the stability, the statistical strength of any given clusters, and the symptom experience of patients who reported symptoms in a cluster ***Design:*** Prospective, longitudinal	**Instrument(s):** MSAS: 32 items ***Analysis:*** EFA with principal components analysis ***Dimension(s):*** Not reported	6 symptom clusters identified: GI: nausea, vomiting, feeling bloated Emotional: with a number of psychological symptoms Respiratory: shortness of breath, cough Hand/foot: numbness; tingling of hand/feet, swelling of arms and legs Body image: hair loss, skin changes, one item "I do not like myself" Nutritional: weight-loss, difficulty swallowing, lack of appetite Change in symptom clusters over time: With slight variations, the six symptom clusters were relatively stable over time	***Strengths:*** First study to explore clusters of symptoms in cancer patients over the first 12 months after diagnosis, reporting cluster patterns, structure, and factor coefficients Used two criterion to evaluate the relationships among symptoms within a cluster (i.e., Cronbach alpha and inter-factor coefficients) Longitudinal study design ***Limitations:*** Did not report the results from the EFA or cluster analysis Patients were primarily of European descent Evaluated patients on a variety of treatment regimens
Skerman et al., 2012[25]	***Instrument(s):***	5 symptom clusters identified:	***Strengths:***

Title: Cancer-related Symptom Clusters for Symptom Management in Outpatients after Commencing Adjuvant Chemotherapy at 6 months and 12 months ***Purpose(s):*** To investigate symptom clusters over time for symptom management of a patient group after commencing adjuvant CTX ***Design:*** Descriptive, longitudinal	Rotterdam Symptom Checklist: 42 items (clinician modified) ***Analysis:*** Common factor analysis with oblique rotation ***Dimension(s):*** Distress	Vasomotor: headache, sweating, hot/cold spells, night sweats, dizziness, numbness/tingling, chest pains, heart pounding/palpitations Oral discomforts: sore throat, sore mouth/pain swallowing, difficulty swallowing, bad taste, loss of taste, dry mouth, deafness UGI: indigestion, heartburn, belching, stomach pain, nausea, low abdominal pain, constipation GI toxicities: poor appetite, vomiting, nausea, shivering, trembling, low abdominal pain, stomach pain, diarrhea, belching, loss of taste, sleepiness, fatigue, weakness Musculoskeletal discomforts/lethargy: weakness, muscle soreness, joint pain, heavy feelings in arms/legs, generalized pain, lower back pains, fatigue, sleepy during day, deafness Change in symptom clusters over time: all five symptom clusters identified were consistent across all three assessments	First longitudinal study to empirically derive symptom clusters Used a symptom inventory with a large number of symptoms Relatively large sample Single cancer treatment Longitudinal study Heterogeneous sample in terms of cancer diagnosis Used sophisticated statistical procedures to determine the number of symptom clusters (i.e. pattern coefficients) and specific symptoms within a cluster (i.e. structure coefficients) ***Limitations:*** Did not assess symptoms prior to treatment Only assessed physical symptoms

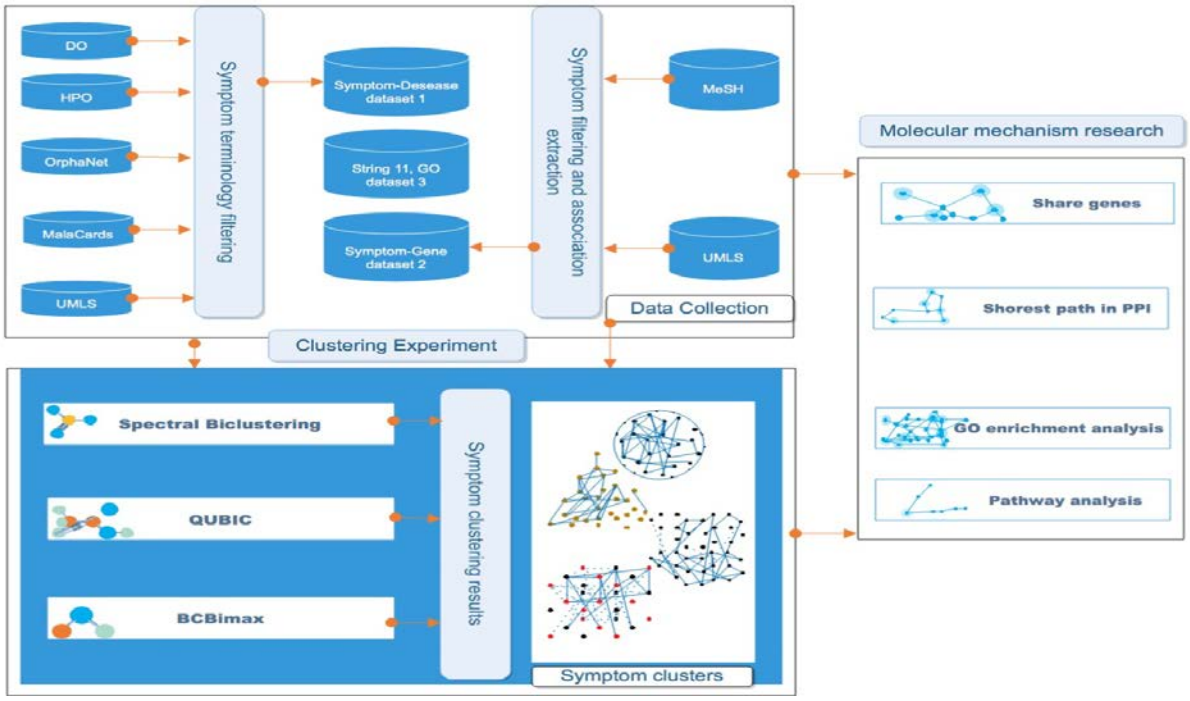
Brown et al., 2011[26] ***Title:*** A Symptom Cluster and Sentinel Symptom Experienced by Women With Lung Cancer ***Purpose(s):*** Describe the occurrence, severity, and clusters of symptoms experienced by women with NSCLC; describe the relationships of demographic and clinical characteristics, health status factors, and meaning of illness with symptom experience and symptom clusters; and determine if a co-occurring sentinel symptom was associated with the presence of symptom clusters ***Design:*** Prospective, longitudinal	***instrument(s):*** Lung Cancer Symptom Scale: 6 items Symptom Query Questionnaire: self-report of symptoms during past 4 weeks obtained during a semi-structured interview Center for Epidemiologic Studies-Depression scale: 20 items Charlson Comorbidity Index: measures the presence of multiple co morbidities ***Analysis:*** Symptoms on the LCSS were coded uniquely and analyzed to determine patterns of co-occurring symptoms ***Dimension(s):*** Occurrence	1 un-named symptom cluster identified: Fatigue, shortness of breath, poor appetite, cough, pain Change in symptom clusters over time: The assessment of symptoms on the previous day revealed a 5-item symptom cluster for 64% of the patients. No predominant symptom cluster was identified for the assessment of the past 4 weeks.	***Strengths:*** First study to evaluate symptom experience of women with early stage NSCLC who were treated surgically Identified a co-occurring sentinel symptom (i.e., pain) that was the most highly correlated symptom with the presence of the 5-symptom cluster ***Limitations:*** Use of SQQ depended on patient's recall of symptoms over the past 4 weeks which may have contributed to under-reporting . Did not address sleep problems, weight loss or decreased concentration reported by 10%–23% of sample Did not use a standard statistical approach to identify symptom clusters Used an instrument with only 6 symptoms to assess symptom clusters

Table 1. the table above show the summary of some earlier works on the same topic which is the foundation of which is research is built

3. MATERIALS & METHODS

The datasets that has been used in this research were taken from different databases as it is depicted on the figure below, different databases have been used as datasets resources needed for symptom clusters analysis. Disease-symptom associations have been searched from the DO[29], HPO[30] and Orphanet[31] databases. We have collected 16383 disease-symptom associations between 13532 diseases and 2378 symptoms from these databases. Symptom-gene we collected 371422 symptoms–gene associations between 2834 symptoms MeSH terms and 17828 genes including 9458 genes from DisGeNet[32]and 8370 Malacard[33], String 11 we have used 841068 and finally, PubMed[18] database used for literature. All these databases have the resources that we need in our research to be complete research. This kind of datasets are similar in two researches[34], [35] .

Figure 1. Conceptual Framework



3.1. Chi-square statistic and control groups for comparison

To show the significance of the related experimental results, we conducted the random procedure as the control groups [45]. In each random procedure, we used Mersenne Twister generator method [46] in python (random package) and reshuffled (10 times) control groups in which have the same amounts as the experimental results. In order to compare experimental results with control groups, we calculated the statistical significance with observed results (e.g., the analysis results of shared genes and ASPLs) by chi-squared test [47], whose formula for calculating a chi-square test is:

$$\sum_{n=1}^{N} \chi^2 = \frac{(O-E)^2}{E} \quad \textbf{(1)}$$

Where O represents the actual count of cases in each selection, E represents the expected value, and X^2 denotes the Chi-square value. In addition, the random experiments are repeated 10 times to guarantee the robustness of random results.

3.2. P-Value

The GO terms shared by the genes in the users in list are compared to the background distribution of the annotation. It is the probability of seeing x or more genes from the input list of n genes annotated to a particular GO term. Given the proportion of genes in the whole genome annotated to that GO term is F out of G. Specifically, hyper geometric distribution is used to calculate the probability of observing at least *x* or more gene from a functional category from an input gene list of size *n* given the background database consists of *G* genes out of which *F* belong to the functional category.

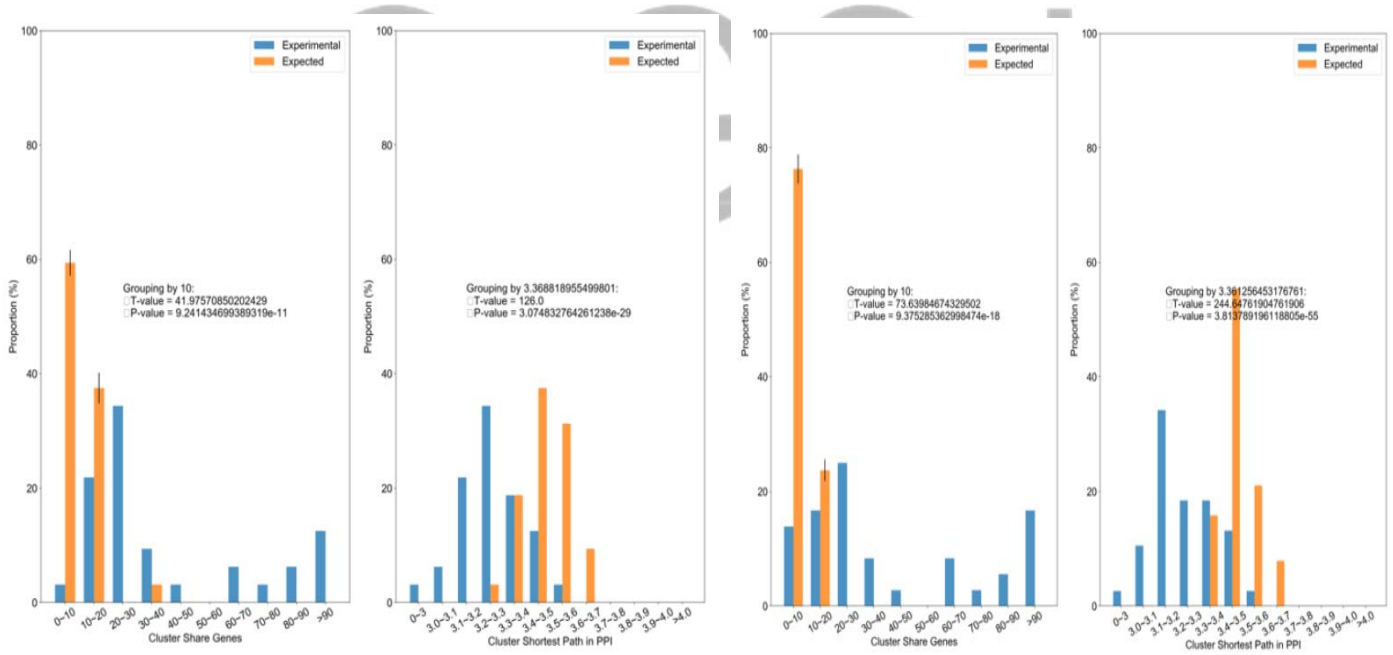
$$\mathrm{p-value} = 1 - \sum_{j=x}^{n} \frac{\binom{F}{j}\binom{G-F}{n-j}}{\binom{G}{n}} \quad (2)$$

It is clear that smaller the p value, more significant is the association of the particular GO term with the group of genes. In our search also we have used p-value of 0.05 as importance level. P-value is the risk of that x huge variety of genes from a Bicluster of dimension X annotated to a unique GO, given P which is the share of genes in the complete genome annotated to that GO term.

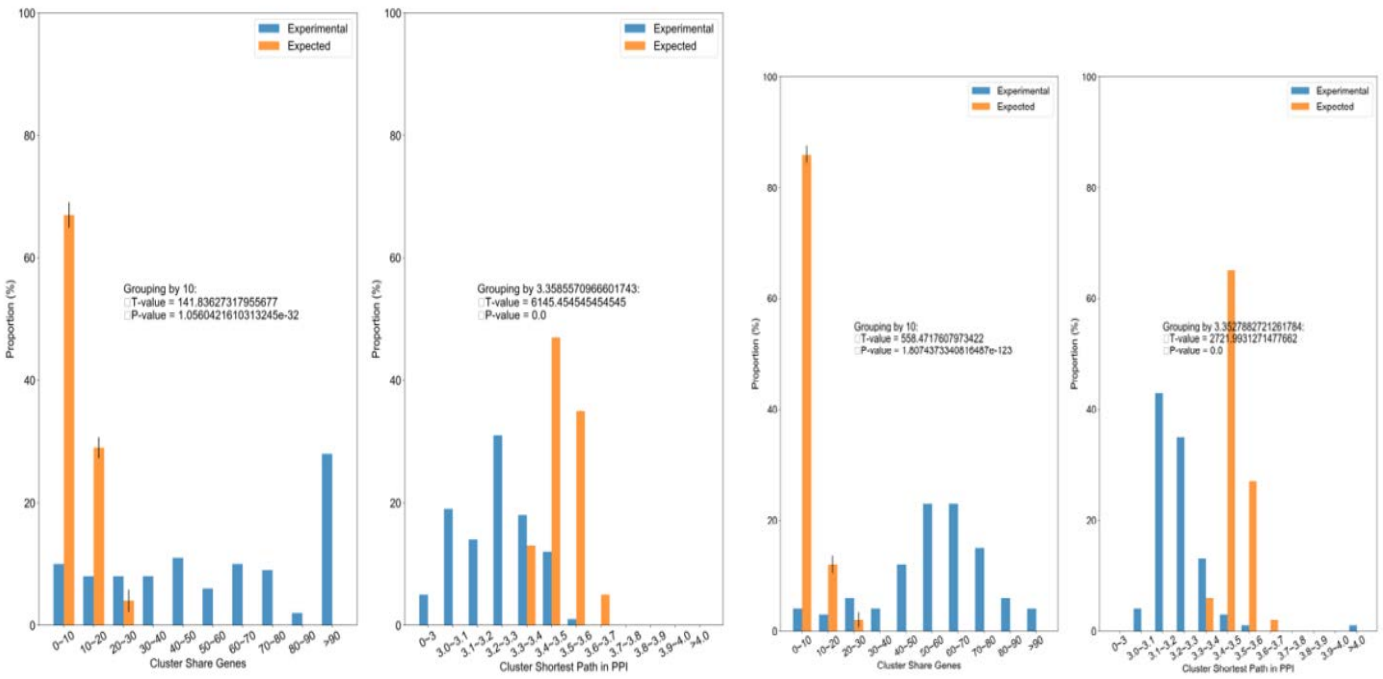
The closer p-value is to zero; the more significant is the affiliation of the specific GO time period with the group of genes.

4. RESULTS

The evaluation is made for comparing the performance of each type of Biclusteing, using symptom genes and string 11 datasets we verify the number of shared genes in the symptom group, and the mean shortest path of symptom group in the PPI network, the difference between symptom group and random symptom group was analyzed, the chi-square test was made for the difference between symptom group and random symptom group under contrast cluster. Finally, according to the average number of shared genes per symptom group and the average shared genes of all symptom groups under random conditions, the ratio of share genes in different range of values is obtained. Finally, chi-square tests were performed on 10 shared genes as the dividing line, as shown in the figure p values are all less than 0.05, which indicates that the result is good. The shortest path method is the same as the shared gene method. By calculating the combination of two genes in two symptoms, we calculate the shortest path in the PPI network under string11 data set as the evaluation index. Again, as with shared gene trials, we used 10 randomly selected symptoms of the same size as a comparison.

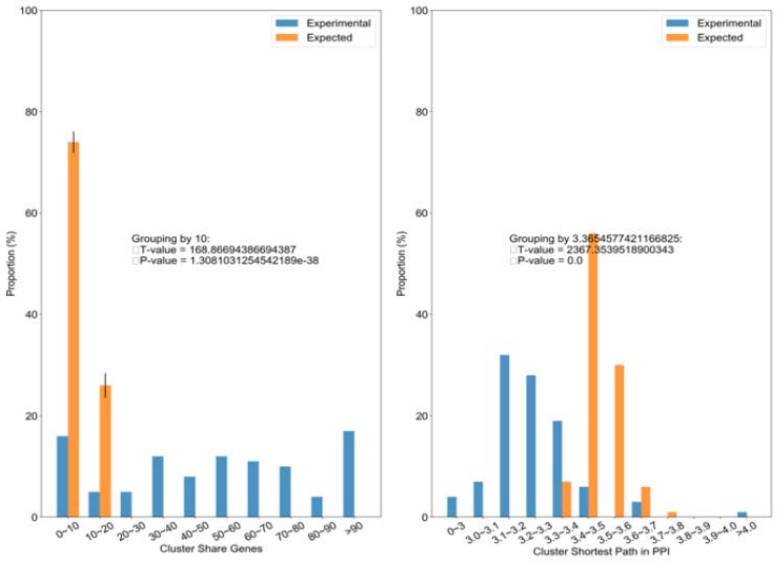


(a) (b)



(c) (d)

(a) Spectral biclustering with 1000 and 150, **(b)** Spectral biclustering with 1000 and 200, **(c)** BIMAX biclustering cluster shared gene and shortest path in PPI result, **(d)** QUBIC biclustering cluster shared gene and shortest path in PPI result when K is 10



(e) QUBIC biclustering cluster shared gene and shortest path in PPI result when K is 30

This table below summarize all calculation computed above in cluster shared genes and cluster shortest path in PPI network where 10 symptoms selected randomly

		Cluster shared genes		**Cluster shortest path in PPI**	
QUBIC	K=10	T-Value	558.471	T-Value	2721.993
		P-value	1.807e-123	P-value	0.0
	K=30	T-Value	168.866	T-Value	2367.35
		P-value	1.308	P-value	0.0
BIMAX	T-Value	T-Value	141.836	T-Value	6145.454
	P-value	P-value	1.056e-32	P-value	0.0
Spectral	(1000,150)	T-Value	41.975	T-Value	126.0
		P-value	9.241e-11	P-value	3.074e-29
	(1000,200)	T-Value	73.639	T-Value	244.647
		P-value	9.375e-11	P-value	3.813e-55

Table 2. The results comparison between three Biclustering algorithms

5. DISCUSSION

In this study we compared three well-established algorithms to evaluate their capabilities of identifying biologically significant groups of co-expressed genes under number conditions. Biclustering experiments by BIMAX, QUBIC and Spectral algorithm based on symptom disease data, based on the molecular mechanism of symptom groups in clustering, we verify the internal criticality of symptom groups, and expect to find new findings in clinical practice. Based on this, we verify the number of shared genes in the symptom group, and the mean shortest path of symptom group in the PPI network, the difference between symptom group and random symptom group was analyzed, the chi-square test was made for the difference. And random symptom group under contrast cluster.

The shared gene evaluation method, we calculated the amount of gene intersection between two symptoms in each symptom group as a symptom group sharing gene by collecting a good curated+ inferred symptom gene

data set, Similarly, for each symptom group, we randomly expected the same number of symptoms as the symptom group as expected data to compare with the results of our cluster of symptoms. Based on this, we used the mean of 10 random trials as the final comparison data for each experiment. Finally, according to the average number of shared genes per symptom group and the average shared genes of all symptom groups under random conditions, the ratio of share genes in different range of values is obtained. Finally, chi-square tests were performed on 10 shared genes as the dividing line, as shown in the figure p values are all less than 0.05, which indicates that the symptoms in specific clusters tend to be closely related to each other with regard to the shared genes or protein-protein interactions. This also means that there exist shared underlying molecular mechanisms between the symptoms in a same cluster.

The shortest path method is the same as the shared gene method. By calculating the combination of two genes in two symptoms, we calculate the shortest path in the PPI network under string11 data set as the evaluation index. Again, as with shared gene trials, we used 10 randomly selected symptoms of the same size as a comparison.

Our results are generally consistent with other surveys of biclustering algorithms. It is similar with the work in [28] [42], we find that Qubic is an effective algorithm that can generate biclusters with high P-value compare with others and also we have based on T-value, as our datasets are huge Qubic is suitable for large size of data. As we use 10 as K we got 100 biclusters for Qubic, spectral found 32 biclusters where Bimax has 100 biclusters.

6. References

[1] A. M. Barsevick, K. Whitmer, L. M. Nail, S. L. Beck, and W. N. Dudley, “Symptom cluster research: Conceptual, design, measurement, and analysis issues,” *J. Pain Symptom Manage.*, vol. 31, no. 1, pp. 85–95, 2006.

[2] S. L. Beck, W. N. Dudley, and A. Barsevick, “Pain, Sleep Disturbance, and Fatigue in Patients With Cancer: Using a Mediation Model to Test a Symptom Cluster,” *Oncol. Nurs. Forum*, vol. 32, no. 3, pp. E48–E55, 2006.

[3] L. Fiorentino, M. Rissling, L. Liu, and S. Ancoli-Israel, “The symptom cluster of sleep, fatigue and depressive symptoms in breast cancer patients: Severity of the problem and treatment options,” *Drug Discov. Today Dis. Model.*, vol. 8, no. 4, pp. 167–173, 2011.

[4] H. Kim, A. M. Barsevick, L. Tulman, and P. A. Mcdermott, “Treatment-Related Symptom Clusters in Breast Cancer : A Secondary Analysis,” *J. Pain Symptom Manage.*, vol. 36, no. 5, pp. 468–479, 2008.

[5] E. Kenne Sarenmalm, M. Browall, and F. Gaston-Johansson, “Symptom burden clusters: A challenge for targeted symptom management. A longitudinal study examining symptom burden clusters in breast cancer,” *J. Pain Symptom Manage.*, 2014.

[6] H. Edwards *et al.*, “Identification of symptom clusters in patients with chronic venous leg ulcers,” *J. Pain Symptom Manage.*, vol. 47, no. 5, pp. 867–875, 2014.

[7] M. J. Dodd, C. Miaskowski, and K. A. Lee, “Occurrence of symptom clusters.,” *J. Natl. Cancer Inst. Monogr.*, no. 32, pp. 76–78, 2004.

[8] A. M. Barsevick, “The Concept of Symptom Cluster,” *Semin. Oncol. Nurs.*, vol. 23, no. 2, pp. 89–98, 2007.

[9] A. Wikman, A. Johar, and P. Lagergren, “Presence of symptom clusters in surgically treated patients with esophageal cancer: Implications for survival,” *Cancer*, vol. 120, no. 2, pp. 286–293, 2014.

[10]C. Ward Sullivan, H. Leutwyler, L. B. Dunn, and C. Miaskowski, “A review of the literature on symptom clusters in studies that included oncology patients receiving primary or adjuvant chemotherapy,” *Journal of Clinical Nursing*. 2018.

[11]M. Dodd *et al.*, “Advancing the science of symptom management. [Review] [49 refs],” *J. Adv. Nurs.*, vol. 33, no. England PT-Journal Article PT-Review PT-Review, Tutorial NO-P30 NR03927 (NINR) LG-English, pp. 668–676, 2001.

[12]N. Karabulu, B. Erci, N. Özer, and S. Özdemir, “Symptom clusters and experiences of patients with cancer,” *J. Adv. Nurs.*, vol. 66, no. 5, pp. 1011–1021, 2010.

[13]S. David, “Population Causes and Consequences of Leading Chronic Diseases: A Comparative Analysis of Prevailing Explanations,” *Milbank Q.*, vol. 86, no. 2, pp. 273–326, 2008.

[14]M. J. Dodd, M. H. Cho, B. A. Cooper, and C. Miaskowski, “The effect of symptom clusters on functional status and quality of life in women with breast cancer,” *Eur. J. Oncol. Nurs.*, vol. 14, no. 2, pp. 101–110, 2010.

[15]M. J. Dodd, C. Miaskowski, and K. A. Lee, “Occurrence of symptom clusters.,” *J. Natl. Cancer Inst. Monogr.*, vol. 0610, no. 32, pp. 76–78, 2004.

[16]C. Miaskowski, B. E. Aouizerat, M. Dodd, and B. Cooper, “Conceptual issues in symptom clusters research and their implications for quality-of-life assessment in patients with cancer.,” *J. Natl. Cancer Inst. Monogr.*, vol. 0610, no. 37, pp. 39–46, 2007.

[17]C. Miaskowski, M. Dodd, and K. Lee, “Symptom clusters: the new frontier in symptom management research.,” *J. Natl. Cancer Inst. Monogr.*, vol. 0610, no. 32, pp. 17–21, 2004.

[18]N. Fiorini, D. J. Lipman, and Z. Lu, "Towards PubMed 2.0," *Elife*, vol. 6, pp. 4–7, 2017.

[19]G. Fan, L. Filipczak, and E. Chow, "Symptom clusters in cancer patients: a review of the literature.," *Curr. Oncol.*, vol. 14, no. 5, pp. 173–9, 2007.

[20]A. Jiménez *et al.*, "Symptom clusters in advanced cancer.," *J. Pain Symptom Manage.*, vol. 42, no. 1, pp. 24–31, 2011.

[21]M. J. Dodd, M. H. Cho, B. A. Cooper, and C. Miaskowski, "The effect of symptom clusters on functional status and quality of life in women with breast cancer," *Eur. J. Oncol. Nurs.*, vol. 14, no. 2, pp. 101–110, 2010.

[22]M. L. Chen and C. C. Lin, "Cancer Symptom Clusters: A Validation Study," *J. Pain Symptom Manage.*, vol. 34, no. 6, pp. 590–599, 2007.

[23]G. Fan, L. Filipczak, and E. Chow, "Symptom clusters in cancer patients: a review of the literature," *Curr. Oncol.*, vol. 14, no. 5, pp. 173–179, 2007.

[24]A. Molassiotis, Y. Wengström, and N. Kearney, "Symptom Cluster Patterns During the First Year After Diagnosis with Cancer," *J. Pain Symptom Manage.*, vol. 39, no. 5, pp. 847–858, 2010.

[25]H. M. Skerman, P. M. Yates, and D. Battistutta, "Cancer-related symptom clusters for symptom management in outpatients after commencing adjuvant chemotherapy, at 6 months, and 12 months," *Support. Care Cancer*, vol. 20, no. 1, pp. 95–105, 2012.

[26]J. K. Brown, M. E. Cooley, C. Chernecky, and L. Sarna, "A Symptom Cluster and Sentinel Symptom Experienced by Women With Lung Cancer," *Oncol. Nurs. Forum*, vol. 38, no. 6, pp. E425–E435, 2011.

[27]C. Miaskowski, "Future Directions in Symptom Cluster Research," *Seminars in Oncology Nursing*. 2016.

[28]L. Li, Y. Guo, W. Wu, Y. Shi, J. Cheng, and S. Tao, "A comparison and evaluation of five biclustering algorithms by quantifying goodness of biclusters for gene expression data," *BioData Min.*, vol. 5, no. 1, pp. 1–10, 2012.

[29]W. A. Kibbe *et al.*, "Disease Ontology 2015 update: An expanded and updated database of Human diseases for linking biomedical knowledge through disease data," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D1071–D1078, 2015.

[30]S. Köhler *et al.*, "The human phenotype ontology in 2017," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D865–D876, 2017.

[31]A. Rath, A. Olry, F. Dhombres, M. M. Brandt, B. Urbero, and S. Ayme, "Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users," *Hum. Mutat.*, vol. 33, no. 5, pp. 803–808, 2012.

[32]J. Piñero *et al.*, "DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes," *Database*, vol. 2015, pp. 1–17, 2015.
[33]N. Rappaport *et al.*, "MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D877–D887, 2017.
[34]K. Yang *et al.*, "Heterogeneous network embedding for identifying symptom candidate genes," *J. Am. Med. Informatics Assoc.*, vol. 25, no. 11, pp. 1452–1459, 2018.
[35]X. Zhou, J. Menche, A. L. Barabási, and A. Sharma, "Human symptoms-disease network," *Nat. Commun.*, vol. 5, no. May, 2014.
[36]P. Dayan, "Unsupervised Learning," *Encycl. Neurosci.*, pp. 4154–4154, 2008.
[37]J. A. Hartigan, "of a Data Matrix Direct Clustering," *Matrix*, vol. 67, no. 337, pp. 123–129, 2011.
[38]G. M. Church, "Biclustering of Expression Data v (I , J)," pp. 93–103, 2000.
[39]A. L. Oliveira and S. C. Madeira, "Biclustering Algorithms for Biological Data Analysis: A Survey," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 1, no. 1, pp. 24–45, 2004.
[40]X. Gan, A. W. C. Liew, and H. Yan, "Discovering biclusters in gene expression data based on high-dimensional linear geometries," *BMC Bioinformatics*, vol. 9, pp. 1–15, 2008.
[41]Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, "Spectral biclustering of microarray data: Coclustering genes and conditions," *Genome Res.*, vol. 13, no. 4, pp. 703–716, 2003.
[42]G. Li, Q. Ma, H. Tang, A. H. Paterson, and Y. Xu, "QUBIC: A qualitative biclustering algorithm for analyses of gene expression data," *Nucleic Acids Res.*, vol. 37, no. 15, 2009.
[43]B. Peter *et al.*, "Comparison of Biclustering Methods : A Systematic Comparison and Evaluation of Biclustering Methods for Gene Expression Data," *Bioinformatics*, vol. 22, no. 9, pp. 1122–29, 2006.
[44]D. A. Hosack *et al.*, "DAVID: Database for Annotation, Visualization, and Integrated Discovery.," *Genome Biol.*, vol. 4, no. 5, p. P3, 2003.
[45] Rescorla R A. Pavlovian conditioning and its proper control procedures[J]. Psychological review, 1967, 74(1): 71.
[46] Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator[J]. ACM Transactions on Modeling and Computer Simulation (TOMACS), 1998, 8(1): 3-30.
[47] Mchugh M L. The chi-square test of independence[J]. Biochemia medica: Biochemia medica, 2013, 23(2): 143-149