

RESEARCH

Open Access



Large-scale mining disease comorbidity relationships from post-market drug adverse events surveillance data

Chunlei Zheng and Rong Xu*

From The International Conference on Intelligent Biology and Medicine (ICIBM) 2018
Los Angeles, CA, USA. 10-12 June 2018

Abstract

Background: Systems approaches in studying disease relationship have wide applications in biomedical discovery, such as disease mechanism understanding and drug discovery. The FDA Adverse Event Reporting System (FAERS) contains rich information about patient diseases, medications, drug adverse events and demographics of 17 million case reports. Here, we explored this data resource to mine disease comorbidity relationships using association rule mining algorithm and constructed a disease comorbidity network.

Results: We constructed a disease comorbidity network with 1059 disease nodes and 12,608 edges using association rule mining of FAERS (14,157 rules). We evaluated the performance of comorbidity mining from FAERS using known disease comorbidities of multiple sclerosis (MS), psoriasis and obesity that represent rare, moderate and common disease respectively. Comorbidities of MS, obesity and psoriasis obtained from our network achieved precisions of 58.6%, 73.7%, 56.2% and recalls 87.5%, 69.2% and 72.7% separately. We performed comparative analysis of the disease comorbidity network with disease semantic network, disease genetic network and disease treatment network. We showed that (1) disease comorbidity clusters exhibit significantly higher semantic similarity than random network (0.18 vs 0.10); (2) disease comorbidity clusters share significantly more genes (0.46 vs 0.06); and (3) disease comorbidity clusters share significantly more drugs (0.64 vs 0.17). Finally, we demonstrated that the disease comorbidity network has potential in uncovering novel disease relationships using asthma as a case study.

Conclusions: Our study presented the first comprehensive attempt to build a disease comorbidity network from FDA Adverse Event Reporting System. This network shows well correlated with disease semantic similarity, disease genetics and disease treatment, which has great potential in disease genetics prediction and drug discovery.

Keywords: FAERS, Association rule mining, Disease comorbidity network

Background

Analysis of disease relationship plays an important role in understanding disease biology and discovering new drug treatments [1]. Disease similarity often indicates underlying genetic connections. For example, genetic loci for complex diseases have been identified by examining

the association between Mendelian diseases and complex diseases [2]. Studying disease relationship is an important strategy for disease gene discovery [3, 4]. Disease relationship is also widely used in drug discovery. For example, discovery of shared genetics of psoriasis and multiple sclerosis led to dimethyl fumarate, an anti-psoriasis drug, to be used in treatment of relapsing-remitting multiple sclerosis [5]. In addition, many drug repurposing strategies are based on disease relationship [6].

*Correspondence: rx@case.edu

¹Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, 2103 Cornell Road, 44106 Cleveland, OH USA



Several disease relationship networks have been reported. Some of them are based on phenotypic similarity, such as disease manifestation network (DMN) [7], clinical trial [8], and some are genetics based, such as human disease network (HDN) derived from Online Mendelian Inheritance in Man (OMIM) [9] and complex disease network (CDN) derived from genome-wide association studies (GWAS) [10]. In addition, comorbidity-based disease networks have also been constructed. Rhetsky et al. developed a statistical model to estimate the co-occurrence relationship for each pair of 160 diseases and demonstrated that comorbidities are genetically linked [11]. Park et al. and Hidalgo et al. detected the comorbidity pairs from the Medicare claims with statistical measures [12, 13]. Roque et al. mined pairwise disease correlations using similar measures from medical records of a psychiatric hospital [14]. All these networks provided valuable information about disease relationship. However, these studies have limitations. For example, Park's study focused on elder patients aged 65 years or older and patients in Roque's study were only from a single health center.

FDA Adverse Event Reporting System (FAERS) contains adverse event reports from manufacturers, consumers and healthcare professionals for all marketed drug and therapeutic biologic products [15]. FAERS is a large-scale database that contains patient diseases, medications, drug adverse events, demographics of around 17 million case reports, which has been extensively explored for detecting drug safety issues. But the rich information of FAERS is still not systematically mined for disease comorbidity. Recently, we used association rule mining algorithm to reveal the link of colorectal cancer with obesity, which demonstrated the feasibility of this method in disease comorbidity study [16, 17]. Multiple disease comorbidities are common in clinic setting [18] and the advantage of association rule mining is that it can flexibly detect multiple disease comorbidity [17]. In this study, we use association rule mining of FAERS to obtain the disease co-occurrence patterns and then constructed a disease comorbidity network (DCN) based on mined association rules.

To our best knowledge, this is the first comprehensive effort in constructing a large-scale disease comorbidity network from 17 million case reports available in FAERS. Through comparative analysis, we demonstrated that disease comorbidity network accurately captures disease-disease relationships published in the literature and has great potential in both disease genetics prediction and drug discovery.

Methods

The overall approach in this study includes three steps (Fig. 1). We firstly used association rule mining of FAERS

indication data to obtain disease comorbidity patterns. Secondly, we constructed a disease comorbidity network (DCN) based on mined rules. Thirdly, we detected the inherent clusters of DCN and examined its correlation with disease genetics, semantic similarity and drug treatment.

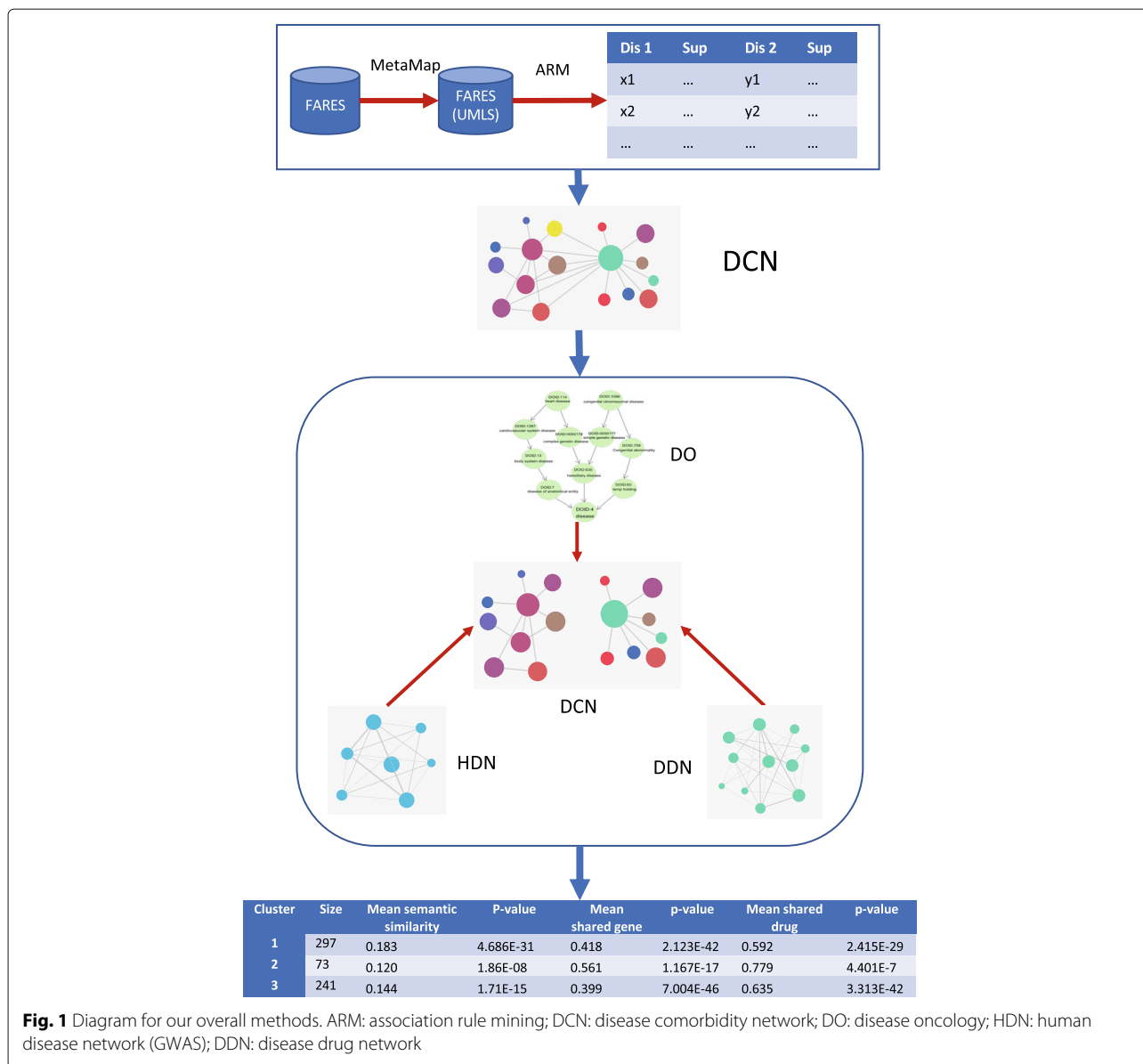
Datasets

FAERS data was downloaded from US Food and Drug Administration (FDA) [15], which contains 17,305,542 case reports for indications from 2004 to 2017. Disease genetic data were extracted from US National Human Genome Research Institute (NHGRI) [19]. The GWAS catalog contains 34,790 disease-gene associations for 14,062 genes and 1665 common complex diseases/traits. Drug-disease associations were extracted from biomedical literature [20–22], which contains 9216 drug-disease pairs for 1483 drugs and 1381 diseases. Disease ontology was downloaded from The OBO Foundry [23], which contains 10,903 disease terms.

Construction of a disease comorbidity network (DCN) by association rule mining of FAERS

FAERS from 2014 to 2017 was used in this study to explore disease comorbidity patterns. After removing reports with unknown indications, the data contained 6,480,372 case reports and represented 15,721 indications of drugs. In order to facilitate downstream analysis, we mapped indications (represented as MedDRA terms) into 12 semantic types that are classified into disorders in UMLS using MetaMap (2016 V2 release) [24]. 12 semantic types of disorders include Acquired Abnormality, Anatomical Abnormality, Cell or Molecular Dysfunction, Congenital Abnormality, Disease or Syndrome, Experimental Model of Disease, Finding, Injury or Poisoning, Mental or Behavioral Dysfunction, Neoplastic Process, Pathologic Function and Sign or Symptom. 12,225 of 15,721 (77.76%) indications were mapped. The clean data contained 6211 diseases and 5,784,501 case reports.

We then applied Frequent Pattern (FP)-growth algorithm (implemented in Weka) [25, 26] into this data to obtain the disease co-occurrence patterns, which is a list of rules between two sets of diseases, represented in the form $\{X \rightarrow Y\}$, for example, $\{\text{anxiety, diabetes mellitus} \rightarrow \text{multiple sclerosis}\}$. FP-growth is a widely used association rule mining algorithm based on FP-tree data structure. Choosing proper support and confidence is a trade-off between precision and recall of disease comorbidities. Here, support > 12 and confidence > 0.5 were used according to performance of validation diseases. Total 14,157 rules were obtained. We constructed an undirected and unweighted disease comorbidity network based on these rules, in which nodes are all diseases at both sides and edges are established between each pair of diseases in both sides.



Clustering of DCN

We used Girvan-Newman algorithm [27, 28] to detect communities in this disease comorbidity network. Girvan-Newman algorithm is based on edge betweenness and edge with biggest betweenness is removed in each iteration. Number of communities of a network depends on how many edges are removed. Modularity metric is computed in each iteration and optimized communities are obtained by maximizing the modularity of network [27].

Correlation analysis of disease comorbidity network with disease genetic network

We constructed a weighted human genetic network (HDN) based on genome wide association data, in which

diseases are represented as nodes and edge is added if two diseases share common genes. Edge weight represents number of share genes between them. Based on genetic information in HDN, we firstly calculated pairwise shared genes in each cluster of DCN obtained from community detection (see above section). Shared genes of a cluster $C(G)$ is defined as average of pairwise shared genes in Eq. 1:

$$C(G) = \frac{1}{m} \sum_{d1 \neq d2, d1, d2 \in D} g(d1, d2) \tag{1}$$

where $d1, d2$ are pairwise diseases in cluster, $g(d1, d2)$ is shared genes between $d1$ and $d2$, D is disease node set in

each cluster and m is number of total pairwise diseases in each cluster.

Correlation analysis of disease comorbidity network with disease treatment network

For computation of shared drugs in each cluster of DCN, we firstly constructed of a disease drug network (DDN) based on FDA drug label data and biomedical literature. Then pairwise shared drugs in a cluster were calculated based on this DDN and share drugs of a cluster $C(D)$ is defined as the average of pairwise shared drugs in Eq. 2:

$$C(D) = \frac{1}{m} \sum_{d1 \neq d2, d1, d2 \in D} d(d1, d2) \quad (2)$$

where $d1, d2$ are pairwise diseases in a cluster, $d(d1, d2)$ is shared drugs between $d1$ and $d2$, D is disease node set in each cluster and m is number of total pairwise diseases in each cluster.

Correlation analysis of disease comorbidity network with disease semantic network

Disease ontology was used for computing semantic similarity of pairwise disease in each cluster of DCN, which is defined as:

$$sim(d1, d2) = \max_{a \in A(d1, d2)} -\log p(a)$$

where $A(d1, d2)$ is the set of common ancestors for $d1$ and $d2$. Semantic similarity of a cluster $C(SIM)$ is computed as average of pairwise disease semantic similarity in Eq. 3:

$$C(SIM) = \frac{1}{m} \sum_{d1 \neq d2, d1, d2 \in D} sim(d1, d2) \quad (3)$$

where $d1, d2$ are pairwise diseases in a cluster, $sim(d1, d2)$ is semantic similarity between $d1$ and $d2$, D is disease node set in each cluster and m is number of total pairwise diseases in each cluster.

Random network was built as a network with the same network structure but nodes are randomly shuffled. We generated 100 random networks and computed the shared genes, shared drugs and semantic similarity in each cluster for each network. T-test was used to compute the significance of each cluster compared with corresponding random networks.

Prioritization of diseases associated with Asthma

In our case study, we used asthma as the seed and random walk with restart (RWR) to rank the diseases that associated with Asthma in DCN. RWR is a ranking algorithm

that has been used to prioritize disease genes [29]. Ranking result is expressed as a probability vector at steady state, representing the probability of each node can be reached from the seed. Assuming p_0 is the seed vector, p_{k+1} , the probability vector at $k + 1$ step, is defined in Eq. 4:

$$p_{k+1} = (1 - \gamma)Mp_k + \gamma p_0 \quad (4)$$

where γ is the restart probability rate and M is adjacency matrix of DCN. γ is set to 0.15. Loop stopped when $|p_{k+1} - p_k| < 10^{-6}$, indicating probability vector is stable.

Results

Properties of Disease Comorbidity Network

Based on association rules from large-scale mining of FAERS, we constructed a disease comorbidity network (DCN), which contains 1059 nodes and 12,608 edges (Fig. 2a). This network is relatively sparse with density of 0.023 (Fig. 2b). Node degrees, i.e., comorbidities for each disease, range from 1 to 685, with median of 7 (Fig. 2c). The nodes with large degrees represent common comorbidities that co-occur with many diseases, such as hypertension, diabetes, depression, anxiety, etc.

Using Girvan-Newman community detection algorithm, DCN can be grouped into 6 clusters with more than 10 nodes (Fig. 2a). We further classified each disease into one of 27 system organ classes (SOC) based on Medical Dictionary for Regulatory Activities (MedDRA), represented as different node colors in the graph. To verify if these clusters reflect inherent disease associations, we computed the top enriched diseases in each cluster (Table 1). We can see that each cluster represents the specific types of diseases. We also noticed that the different types of diseases are grouped together, which reflects the additional level of disease associations.

Evaluate disease comorbidity mining using three diseases: multiple sclerosis, psoriasis and obesity

Disease comorbidity is a complicated and dynamic concept and no systematic database is available, which makes evaluation of our disease comorbidity network difficult. Here, we evaluated DCN by comparison with literature reports, especially using information from corresponding national health organizations. Three diseases, including multiple sclerosis (MS), psoriasis and obesity were chosen for this purpose, which represent rare, moderate and common disease separately. Multiple sclerosis (MS) is a demyelinating disease in which the insulating covers of nerve cells in the brain and spinal cord are damaged. It is estimated that 2.3 million people have

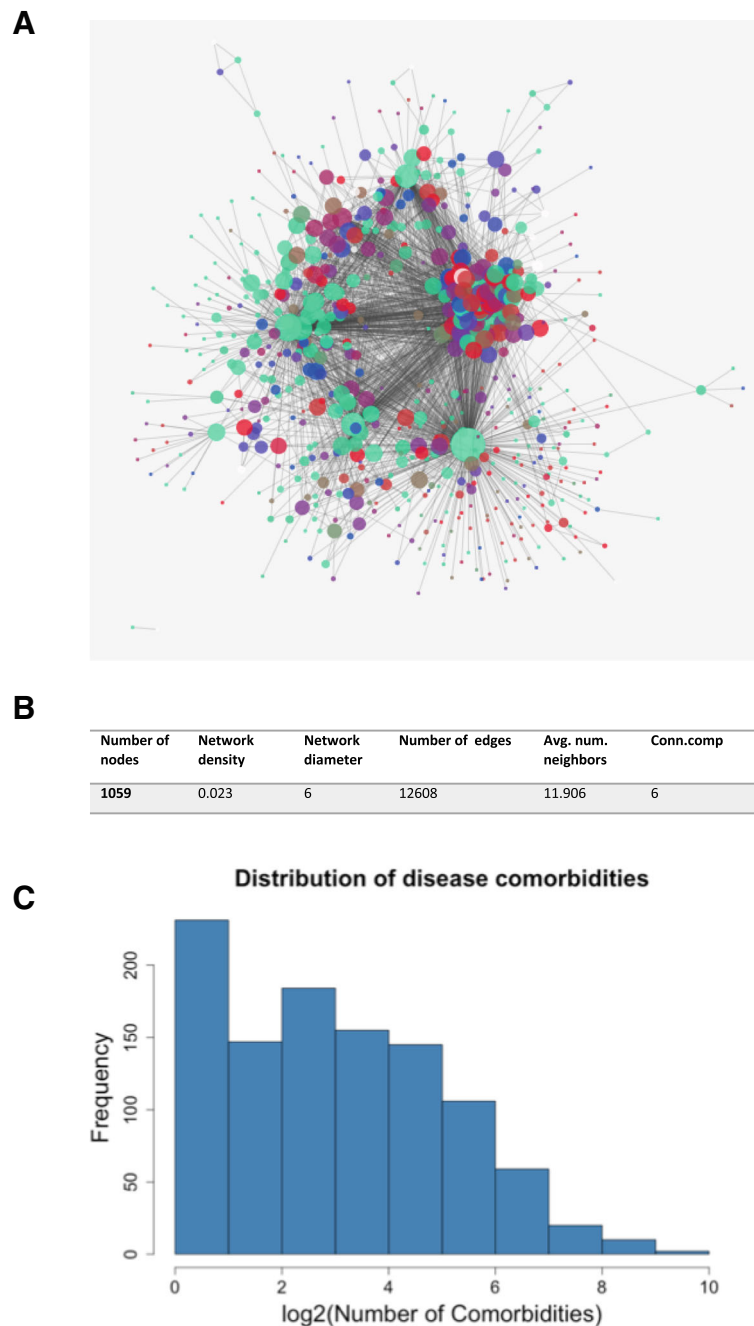


Fig. 2 Characteristics of disease comorbidity network. **a** Visualization of DCN. Diseases are represented as nodes and the size of each node is proportional to the degree of that node. Node color represents disorder class (SOC in MedDRA) to which it belongs. Edges between nodes are represented as the co-occurrence of diseases. **b** Basic network property of DCN. **c** Distribution of disease comorbidity in DCN

MS worldwide. Psoriasis is a chronic inflammatory disorder associated with significant morbidity and mortality. The prevalence of psoriasis among US adults ages 20 years and older is 3.2% [30]. Obesity is a medical condition in which excess body fat has accumulated to the extent that it may have a negative effect on health. More than one-third (36.5%) of U.S. adults have obesity [31].

Table 2 lists comorbidities of MS, psoriasis and obesity. Comorbidities of MS were extracted from National Multiple Sclerosis Society [32] and a literature report [33]; Comorbidities of psoriasis were obtained from National Psoriasis Foundation [34]; Comorbidities of obesity were obtained from Centers for Disease Control and Prevention [35].

Table 1 Top enriched diseases in each cluster

Cluster	Enriched diseases	Folds	P value
1	Cardiac disorders	3.72	3.08e-07
	Vascular disorders	2.28	1.46e-04
2	Metabolism and nutrition disorders	3.17	3.11e-05
3	Psychiatric disorders	8.71	3.33e-33
4	Renal and urinary disorders	2.63	2.06e-02
	Musculoskeletal and connective tissue disorders	2.54	1.48e-02
5	Eye disorders	10.38	3.56e-03
6	Infections and infestations	5.19	2.57e-03

In DCN network, we considered all its neighbor nodes as comorbidities of given disease. To test the performance of our network, we extracted the comorbidities of these three diseases from DCN and compared with literature report as mentioned above. DCN achieved precision of 58.6%, 73.7% and 56.2%, and recall of 87.5%, 69.2% and 72.7% for MS, obesity and psoriasis separately (Fig. 3).

Disease comorbidity network significantly correlates with disease semantic similarity

Semantic similarity is a measurement that calculates the disease distance based on disease ontology. High semantic similarity between two diseases indicates that

Table 2 Comorbidities for obesity, psoriasis and multiple sclerosis

Disease	Comorbidities
Obesity	Hypertension, Coronary heart disease, Stroke, Dyslipidemia, Type 2 diabetes,
	Gallbladder disease, Osteoarthritis, Sleep apnea and breathing problems,
	Some cancers (endometrial, breast, colon, kidney, gallbladder and liver),
	Mental illness (depression, anxiety)
Psoriasis	Cardiac event, Stroke, Crohn's Disease, Diabetes, Metabolic syndrome, Obesity,
	Osteoporosis, Uveitis and Liver Disease, Cancer, Depression
Multiple Sclerosis	Spasticity, Bladder disorder, Bowel problem, Vision problem, Fatigue, Weakness,
	Chronic lung disease, Hypertension, Diabetes, Numbness, Sexual problem, Pain,
	Headache, Epilepsy, Cognitive problem, Seizure, Tremor,
	Psychiatric problems (depression, bipolar disorder, anxiety and schizophrenia)

they share more pathological processes. Pairwise disease semantic similarity in each cluster was computed as mentioned in method and cluster semantic similarity is the average of all pairwise disease similarities in that cluster. Compared with random networks, semantic similarity in each cluster of DCN is significantly higher (Table 3).

Disease comorbidity network significantly correlates with disease genetics

Disease comorbidities often have common genetic causes and common phenotypic features. To test if our network essentially captures this observation, we computed the shared genes in each cluster and compared with random networks. Shared genes in pairwise diseases were obtained from disease-gene network constructed using GWAS resources. Overall shared genes in each cluster is the average of shared genes in all disease pairs in that cluster. The result showed that each cluster of DCN shares significantly more genes than that in random networks, indicating that DCN captures genetic relationship between diseases (Table 4).

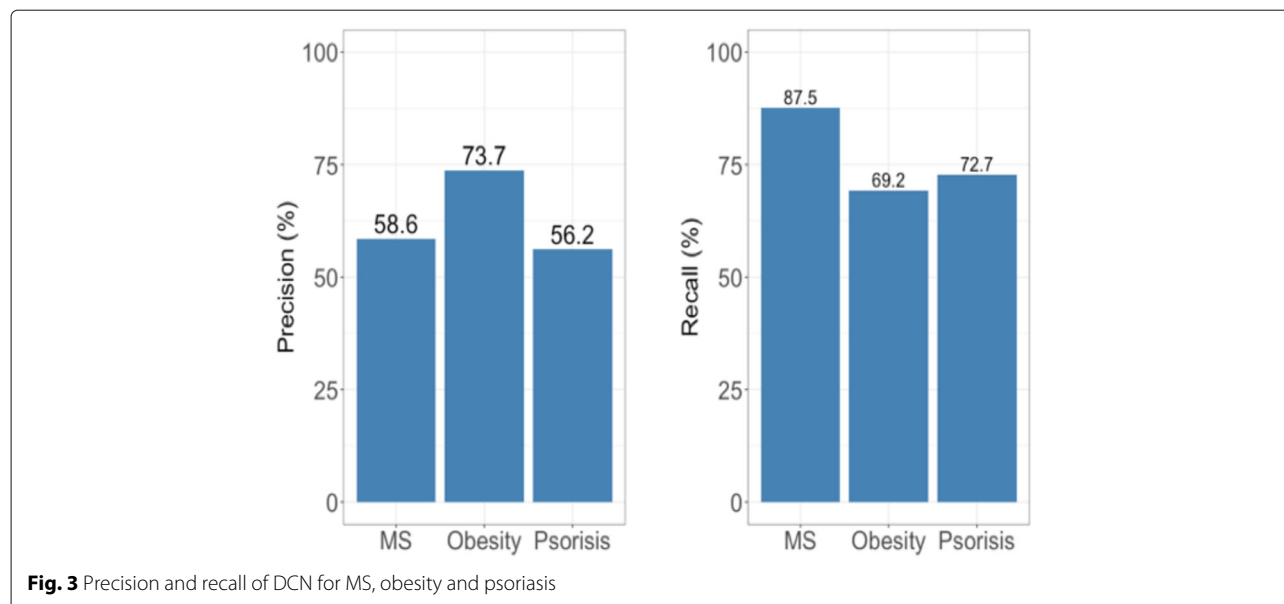
Disease comorbidity network significantly correlates with disease drug treatment

Furthermore, we investigated if DCN captures disease drug treatment information. Similar method was used to compute shared drugs in each disease comorbidity cluster, we used disease-drug network to obtain shared drugs between two diseases. Compared with random networks, shared drugs in each cluster are also significantly higher (Table 5). Taken together, these correlation analyses demonstrate that our disease comorbidity network essentially captures disease-disease relationship.

Disease comorbidity network reveals interesting disease associations/comorbidities - a case study

To demonstrate that DCN can be used for discovering novel disease relationship, we use asthma as an example. Asthma is a common long-term inflammatory disease of the airways of the lungs. Asthma is still an incurable disease. Main purpose of current medical treatments is to control symptoms. Many comorbidities of asthma have been observed, such as hay fever, allergy, obesity, sleep apnea, anxiety, depression, chest pain and cough, which makes asthma more difficult to control. Nevertheless, some diseases, such as hypertension and cardiovascular disease, shows association with asthma, but is still in controversial [36–38]. Therefore, understanding comorbidities of asthma is important for disease management and underlying biology.

We used random walk with restart to find diseases that are associated with asthma. Asthma was used as



the seed and we ranked the other diseases in DCN, which represents the probabilities of each disease can be reached. We expected that comorbidities of asthma should rank high since they generally share common genetics. We can see that all of them rank in top 7% except obesity (Table 6), which further demonstrated the robust and usefulness of our network. Interestingly, hypertension and cardiovascular diseases, including atrial fibrillation, cardiac failure and stroke, also ranked very high (Table 6), suggesting that they might be also closely related to asthma. Indeed, two recent studies supports our observation. One is a large sleep cohort study that demonstrated that late-onset asthma significantly increases cardiovascular diseases [39]. Another is a case-control study from Kaiser Permanente database that established hypertension is a comorbidity of asthma [40]. These evidences strongly demonstrate that our disease comorbidity

network is able to reliably capture disease-disease relationship and have great potential to reveal novel disease relationships.

Discussion

We constructed a disease comorbidity network by association rule mining of large-scale post market surveillance database. This network is able to accurately capture disease comorbidities of literature reported and well correlated with disease semantic similarity, disease genetics and drug treatments. More interestingly, it can also discover associated diseases in debate due to inconsistent reports in literature, which can be explained by large-scale of FAERS dataset. All these properties of DCN indicate that it has a great potential to be used in disease genetics prediction and drug discovery. We note that the way FAERS data is collected may bias the network as compared to the general population. Further evaluation of these biases will be required in subsequent work.

Table 3 Statistics for semantic similarity of disease comorbidity network

Cluster	Size	Semantic similarity (DCN)	Mean semantic similarity (Random)	p-value
1	297	0.183	0.109	4.686E-31
2	73	0.120	0.095	1.860E-08
3	241	0.144	0.109	1.710E-15
4	335	0.131	0.110	2.048E-10
5	63	0.322	0.085	9.080E-92
6	50	0.200	0.097	5.273E-65

Table 4 Statistics for shared genes of disease comorbidity network

Cluster	Size	Shared gene (DCN)	Mean shared gene (Random)	p-value
1	297	0.418	0.061	2.123E-42
2	73	0.561	0.012	1.167E-17
3	241	0.399	0.015	7.004E-46
4	335	0.433	0.053	5.733E-34
5	63	0.534	0.009	3.520E-18
6	50	0.413	0.029	1.836E-18

Table 5 Statistics for shared drugs of disease comorbidity network

Cluster	Size	Shared drug (DCN)	Mean shared drug (Random)	p-value
1	297	0.592	0.136	2.415E-29
2	73	0.779	0.178	4.401E-7
3	241	0.635	0.062	3.313E-42
4	335	0.542	0.070	1.999E-36
5	63	0.610	0.180	4.149E-6
6	50	0.678	0.110	2.394E-5

Previously, we also constructed a disease manifestation network (DMN) based on UMLS [7], which has been successfully used for disease gene prediction and drug discovery [17]. In addition, other disease-disease network based on electronic health record [14] and Medicare [12] are also available. Different data sources contain redundant and complementary information about disease relationship. We prospect that integration of these data sources may provide more power in disease gene discovery and drug repurposing.

Conclusions

We built a comprehensive disease comorbidity network from FAERS using association rule mining. This network not only effectively retrieves known comorbidities of given disease, but is capable to reveal new disease-disease associations. Additionally, correlation analysis also shows that it reflects inherent disease relationships. This work provides a new source for study of disease genetics and drug discovery.

Table 6 Ranks of asthma associated diseases from disease comorbidity network

Disease	Rank (%)	Comorbidity
Hay fever (Allergic rhinitis)	2.55	Yes
Allergy	4.25	Yes
Obesity	22.10	Yes
Obstructive sleep apnea	6.61	Yes
Anxiety	0.66	Yes
depression	0.57	Yes
chest pain	1.79	Yes
cough	3.68	Yes
hypertension	0.00	No
atrial fibrillation	0.09	No
cardiac failure	1.51	No
stroke	2.17	No

Abbreviations

DCN: Disease comorbidity network; DDN: Disease drug network; FAERS: FDA adverse event reporting system; FP: Frequent pattern; GWAS: Genome-wide association studies; HDN: Human disease network; MedDRA: Medical dictionary for regulatory activities; MS: Multiple sclerosis; OMIM: Online mendelian inheritance in man; RWR: Random walk with restart; SOC: System organ classes

Acknowledgements

Not applicable.

Funding

This work was supported by the NIH Director's New Innovator Award under the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (DP2HD084068, Xu), NIH National Institute of Aging (R01 AG057557-01, R01 AG061388-01, R56 AG062272-01, Xu), American Cancer Society Research Scholar Grant (RSG-16-049-01 - MPC, Xu) and NIH Clinical and Translational Science Collaborative of Cleveland (1UL1TR002548-01, Konstan). The funding body played no role in the design of the study, interpretation of data and writing the manuscript. Publication of this article was sponsored by the NIH Director's New Innovator Award under the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (DP2HD084068, Xu).

Availability of data and materials

The dataset supporting the conclusions of this article is available in the http://nlp.case.edu/public/data/FAERS_comb

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 19 Supplement 17, 2018: Selected articles from the International Conference on Intelligent Biology and Medicine (ICIBM) 2018: bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-17>.

Authors' contributions

RX conceived and designed this study. CZ implemented the algorithms and prepared the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Published: 28 December 2018

References

- Oti M, Huynen MA, Brunner HG. Phenome connections. *Trends Genet.* 2008;24(3):103–6. <https://doi.org/10.1016/j.tig.2007.12.005>.
- Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabani H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A. A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk. *Cell.* 2013;155(1):70–80. <https://doi.org/10.1016/j.cell.2013.08.030>.
- Zhu C, Wu C, Aronow BJ, Jegga AG. Computational approaches for human disease gene prediction and ranking. *Adv Exp Med Biol.* 2014;799:69–84. https://doi.org/10.1007/978-1-4614-8778-4_4.
- Opap K, Mulder N. Recent advances in predicting gene-disease associations. *F1000Res.* 2017;6:578. <https://doi.org/10.12688/f1000research.10788.1>.
- Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M, Yang M, Raghupathi K, Novas M, Sweetser MT, Vigiotta V, Dawson KT,

- CONFIRM Study Investigators. Placebo-controlled phase 3 study of oral bg-12 or glatiramer in multiple sclerosis. *N Engl J Med.* 2012;367(12):1087–97. <https://doi.org/10.1056/NEJMoa1206328>.
6. Vanhaelen Q, Mamoshina P, Aliper AM, Artemov A, Lezhnina K, Ozerov I, Labat I, Zhavoronkov A. Design of efficient computational workflows for in silico drug repurposing. *Drug Discov Today.* 2017;22(2):210–22. <https://doi.org/10.1016/j.drudis.2016.09.019>.
 7. Chen Y, Zhang X, Zhang G-Q, Xu R. Comparative analysis of a novel disease phenotype network based on clinical manifestations. *J Biomed Inform.* 2015;53:113–20. <https://doi.org/10.1016/j.jbi.2014.09.007>.
 8. Haslam B, Perez-Breva L. Learning disease relationships from clinical drug trials. *J Am Med Inform Assoc.* 2017;24(1):13–23. <https://doi.org/10.1093/jamia/ocw003>.
 9. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. *Proc Natl Acad Sci U S A.* 2007;104(21):8685–90. <https://doi.org/10.1073/pnas.0701361104>.
 10. Barrenas F, Chavali S, Holme P, Mobini R, Benson M. Network properties of complex human disease genes identified through genome-wide association studies. *PLoS ONE.* 2009;4(11):8090. <https://doi.org/10.1371/journal.pone.0008090>.
 11. Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci U S A.* 2007;104(28):11694–9. <https://doi.org/10.1073/pnas.0704820104>.
 12. Park J, Lee D-S, Christakis NA, Barabási A-L. The impact of cellular networks on disease comorbidity. *Mol Syst Biol.* 2009;5:262. <https://doi.org/10.1038/msb.2009.16>.
 13. Hidalgo CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol.* 2009;5(4):1000353.
 14. Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, Søeby K, Bredkjær S, Juul A, Werge T, Jensen LJ, Brunak S. Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Comput Biol.* 2011;7(8):1002141. <https://doi.org/10.1371/journal.pcbi.1002141>.
 15. FAERS. <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm345338.htm>. Accessed 2017.
 16. Chen Y, Xu R. Mining cancer-specific disease comorbidities from a large observational health database. *Cancer Inform.* 2014;13(Suppl 1):37–44. <https://doi.org/10.4137/CIN.S13893>.
 17. Chen Y, Li L, Xu R. Disease comorbidity network guides the detection of molecular evidence for the link between colorectal cancer and obesity. *AMIA Jt Summits Transl Sci Proc.* 2015;2015:201–6.
 18. van den Akker M, Buntinx F, Metsmakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol.* 1998;51(5):367–75.
 19. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, Parkinson H. The nhgri was catalog, a curated resource of snp-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):1001–6. <https://doi.org/10.1093/nar/gkt1229>. Accessed 2017.
 20. Xu R, Wang Q. Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing. *BMC Bioinformatics.* 2013;14:181. <https://doi.org/10.1186/1471-2105-14-181>.
 21. Xu R, Li L, Wang Q. Towards building a disease-phenotype knowledge base: extracting disease-manifestation relationship from literature. *Bioinformatics.* 2013;29(17):2186–94. <https://doi.org/10.1093/bioinformatics/btt359>.
 22. Xu R, Wang Q. Automatic signal extraction, prioritizing and filtering approaches in detecting post-marketing cardiovascular events associated with targeted cancer drugs from the fda adverse event reporting system (faers). *J Biomed Inform.* 2014;47:171–7. <https://doi.org/10.1016/j.jbi.2013.10.008>.
 23. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H, Schriml LM. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* 2015;43(Database issue):1071–8. <https://doi.org/10.1093/nar/gku1011>.
 24. MetaMap (2016 v2 release). <https://doi.org/metamap.nlm.nih.gov>. Accessed 2017.
 25. Frank E, Hall MA, Witten IH. The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”, Fourth ed. San Francisco: Morgan Kaufmann; 2016.
 26. Jiawei Han YY, Pei J. Mining Frequent Patterns Without Candidate Generation. In: Proceedings of the 2000 ACM-SIGMOD International Conference on Management of Data. New York: ACM; 2000.
 27. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004;69(2 Pt 2):026113. <https://doi.org/10.1103/PhysRevE.69.026113>.
 28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.
 29. Li L, Wang Y, An L, Kong X, Huang T. A network-based method using a random walk with restart algorithm and screening tests to identify novel genes associated with menière’s disease. *PLoS ONE.* 2017;12(8):0182592. <https://doi.org/10.1371/journal.pone.0182592>.
 30. Rachakonda TD, Schupp CW, Armstrong AW. Psoriasis prevalence among adults in the united states. *J Am Acad Dermatol.* 2014;70(3):512–6. <https://doi.org/10.1016/j.jaad.2013.11.013>.
 31. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United states, 2011–2014. *NCHS Data Brief.* 2015;219:1–8.
 32. National Multiple Sclerosis Society. <https://www.nationalmssociety.org/Symptoms-Diagnosis/MS-Symptoms>. Accessed 2017.
 33. Marrie RA HH. General health issues in multiple sclerosis: comorbidities, secondary conditions, and health behaviors. *Contin (Minneapolis Minn).* 2013;19(4 Multiple Sclerosis):1046–57.
 34. National Psoriasis Foundation. <https://www.psoriasis.org/about-psoriasis/related-conditions>. Accessed 2017.
 35. Centers for Disease Control and Prevention. <https://www.cdc.gov/healthyweight/effects/index.html>. Accessed 2017.
 36. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease?. *Int J Epidemiol.* 2004;33(4):743–8. <https://doi.org/10.1093/ije/dyh081>.
 37. Schanen JG, Iribarren C, Shahar E, Punjabi NM, Rich SS, Sorlie PD, Folsom AR. Asthma and incident cardiovascular disease: the atherosclerosis risk in communities study. *Thorax.* 2005;60(8):633–8. <https://doi.org/10.1136/thx.2004.026484>.
 38. Iribarren C, Tolstykh IV, Miller MK, Sobel E, Eisner MD. Adult asthma and risk of coronary heart disease, cerebrovascular disease, and heart failure: a prospective study of 2 matched cohorts. *Am J Epidemiol.* 2012;176(11):1014–24. <https://doi.org/10.1093/aje/kws181>.
 39. Tattersall MC, Barnett JH, Korcarz CE, Hagen EW, Peppard PE, Stein JH. Late-onset asthma predicts cardiovascular disease events: The wisconsin sleep cohort. *J Am Heart Assoc.* 2016;5(9):. <https://doi.org/10.1161/JAHA.116.003448>.
 40. Christiansen SC, Schatz M, Yang S-J, Ngor E, Chen W, Zuraw BL. Hypertension and asthma: A comorbid relationship. *J Allergy Clin Immunol Pract.* 2016;4(1):76–81. <https://doi.org/10.1016/j.jaip.2015.07.009>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

