Disease Comorbidity Network Guides the Detection of Molecular Evidence for the Link Between Colorectal Cancer and Obesity

Yang Chen
Division of Medical Informatics,
Department of EECS
Case Western Reserve University
Cleveland, Ohio, USA
yxc233@case.edu

Li Li, MD, PhD
Departments of Family Medicine and
Community Health, Epidemiology and
Biostatistics, Case Western Reserve
University, Cleveland, OH, USA
li.li@uhhospitals.org

Rong Xu, PhD
Division of Medical Informatics,
Case Western Reserve University
Cleveland, Ohio, USA
rxx@case.edu

Abstract—Epidemiological studies suggested that obesity increases the risk of colorectal cancer (CRC). The genetic connection between CRC and obesity is multifactorial and inconclusive. In this study, we hypothesize that the study of shared comorbid diseases between CRC and obesity can offer unique insights into common genetic basis of these two diseases. We constructed a comorbidity network based on mining health data for millions of patients. We developed a novel approach and extracted the diseases that play critical roles in connecting obesity and CRC in the comorbidity network. Our approach was able to prioritize metabolic syndrome and diabetes, which are known to be associated with obesity and CRC through insulin resistance pathways. Interestingly, we found that osteoporosis was highly associated with the connection between obesity and CRC. Through gene expression meta-analysis, we identified novel genes shared among CRC, obesity and osteoporosis. Literature evidences support that these genes may contribute in explaining the genetic overlaps between obesity and CRC.

Keywords—comorbidity network; colorectal cancer; obesity; osteoporosis; association rule mining; gene expression

I. INTRODUCTION

Comorbidity studies often detect unexpected disease links [1] and offer novel insights into the genetic mechanisms of diseases [2, 3]. A number of epidemiological studies suggest that obesity increases the risk of colorectal cancer (CRC) [4-6]. Based on these evidences of co-occurrence, many genetic factors have been proposed to explain the role of obesity in the development of CRC. For example, both animal and human studies have demonstrated that the increased release of insulin and reduced insulin signaling play roles in obesity and colorectal carcinogenesis [7-9]. Experiments also show that obesity leads to altered level of adipocytokines, such as Adiponectin [10-12] and leptin [13, 14], which may either prevent or foster carcinogenesis.

The mechanism for the association between obesity and CRC is multifactorial and inconclusive [6, 15, 16]. Shared comorbidities between obesity and CRC can provide unique insights into the common genetic basis for the two diseases.

For example, type 2 diabetes is highly correlated with obesity and was identified as a risk factor for CRC [17]. A few studies then discovered that genetic factors of insulin resistance, which occur in type 2 diabetes, contribute in explaining the role of obesity in CRC [18]. However, both obesity and CRC are heterogeneous conditions. Over 40% of the obese population is not characterized by the presence of insulin resistance [19]. We hypothesize that systems approaches to studying the diseases that are phenotypically-significant to both CRC and obesity may offer new insights into the common molecular mechanisms between the two interconnected diseases.

Systematic comorbidity studies have been conducted previously, but mostly focused on pairwise comorbidities and their genetic overlaps. Rhetsky et al. developed a statistical model to estimate the co-occurrence relationship for each pair of 160 diseases [20], and demonstrated that comorbidities are genetically linked. Park et al. [21] and Hidalgo et al. [22] detected the comorbidities pairs from the Medicare claims (which only contain senior patients ages 65 or older) with statistical measures. Roque et al. mined pairwise disease correlations using similar measures from medical records of a psychiatric hospital [23]. Recently, we extracted comorbidity patterns from a publically accessible database, which contains disease records for millions of patients at all ages, using an association rule mining approach [24, 25].

In this study, we constructed a disease comorbidity network based on our previous work. We developed a novel approach to detect diseases that have strong connections with both obesity and CRC in the comorbidity network. Specifically, we extracted the local network consisting of all the paths between obesity and CRC, and prioritized the nodes (diseases) that play critical roles in maintaining the connection between the two diseases (Fig.1). Substantial literature evidences can support that the top ranked diseases have associations with both obesity and CRC. We investigated the gene expression profiles of a prioritized comorbid disease to facilitate detecting novel genetic basis underlying the link between obesity and CRC. Our approach is generalizable to study the genetic basis for other disease associations.
II. MATERIALS AND METHODS

Fig. 2 shows the three steps of our approach. We first mined disease comorbidity relationships from large amounts of patient records and constructed a disease comorbidity network. We then extracted the local comorbidity cluster for obesity and CRC and prioritize the candidate comorbidity that plays a critical role in connecting the two diseases. Finally we conducted gene expression meta-analysis to identify common genes shared by obesity, CRC and the prioritized comorbidity.

A. Construct Disease Comorbidity Network

We mined disease comorbidity relationships from the FDA adverse event reporting system. The database contains records (2004-2013) of 3,354,043 patients (male and female at all age levels) and 10,112 disorders. Our previous studies [24, 25] have demonstrated that this database is useful in mining comorbidity patterns among diverse patient populations.

We applied the association rule mining approach to detect disease comorbidity relationships from the patient-disease pairs. Association rule mining can flexibly detect strong co-occurrence relationships among sets of diseases, represented in the form \( x \rightarrow y \). We collected all diseases in the set \( x \) and \( y \) in each pattern, assuming they have comorbidity relationships with each other, and established an edge between each pair of diseases in \( x \cup y \) to construct the comorbidity network [24].

B. Prioritize the Diseases That Have Strong Associations with Both Obesity and CRC

We extracted the local network consisting of the paths from obesity to CRC in the disease comorbidity network. The local network thus includes the nodes that may represent different aspects of the relationship between obesity and CRC. We implemented breadth first search to enumerate the paths, and limited the paths within four steps.

Then we ranked the nodes in the local network, except obesity and CRC, based on how important they are in maintaining the local network structure and the connection between obesity and CRC. We used the degree and betweenness centrality to characterize the importance of each node in the flowing of the network. The degree of a node becomes higher if more paths between obesity and CRC pass through this node. The betweenness evaluates the number of times that the node acts as the bridge along the shortest paths. Removing the nodes with highest degree or betweenness can easily break down the connection between obesity and CRC.
We investigated the top ranked diseases based on both ranking methods, and used the unexpected ones to guide the detection of genetic associations between obesity and CRC.

C. Identify Gene Overlaps Through Gene Expression Meta-analysis

We chose a top ranked disease on the path between obesity and CRC, and then conducted gene expression meta-analysis for the prioritized disease, obesity and CRC, respectively, to detect new genetic explanations for the relationship between obesity and CRC. Gene expression normalized data (SOFT files) were downloaded from NCBI GEO omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) using the R package GEOquery [28]. Then, we performed microarray meta-analyses for each disease independently using the R package MetaDE [29]. MetaDE implements meta-analysis methods for differential expression analysis, and we used the Fisher’s method. Significant differentially expressed genes (DEGs) were selected as those displaying a FDR corrected p-value <0.05. Last, we extracted the common significant genes for the three diseases.

III. Results

A. Local Disease Comorbidity Network Models the Connection Between Obesity and CRC

We extracted 7006 comorbidity association rules with the confidence larger than 50% from the patient records across ten years. The comorbidity network based on these rules contains 771 nodes and 15,667 edges. Fig.3 shows the local network consisting of all the 119 paths (no longer than four steps) from obesity to CRC. A total of 24 nodes in the local network are the candidate diseases, which have associations with both obesity and CRC, and may indicate different aspects of the relationship between the two diseases.

B. Osteoporosis Shows High Comorbidity Associations with Both CRC and Obesity

Table 1 shows the top five nodes sorted by degree and betweenness in the local network. In either way of ranking, hypertension, diabetes and hyperlipaemia were in top three and closely related with both obesity and CRC. Substantial literature evidences support that the metabolic syndrome components, hypertension and hyperlipaemia, as well as diabetes have association with obesity and CRC through insulin resistance in substantial literature [6-9, 18]. These three disorders also independently increase the risk of CRC and colorectal adenoma [6, 17, 18]. The top ranked comorbidities demonstrated the validity of our network analysis approach.

Significantly, osteoporosis was ranked highly by both

| Rank | Ranked by degree | Ranked by betweenness |
|------|------------------|-----------------------|
| 1    | Hypertension     | Degree                |
| 2    | Diabetes mellitus| Degree                |
| 3    | Hyperlipaemia    | Degree                |
| 4    | Osteoporosis     | Degree                |
| 5    | Hypothyroid      | Degree                |
|      |                  | Nodes                 |
|      |                  | Betweenness           |
| 1    | Hypertension     | 26                    |
| 2    | Diabetes mellitus| 24                    |
| 3    | Hyperlipaemia    | 22                    |
| 4    | Osteoporosis     | 14                    |
| 5    | Hypothyroid      | 14                    |
|      |                  | 60.2                  |
|      |                  | 55.9                  |
|      |                  | 35.2                  |
|      |                  | 12.3                  |
|      |                  | 9.5                   |

Significantly, osteoporosis was ranked highly by both
TABLE II. COMMON GENES SHARED BY OBESITY, CRC AND OSTEOPOROSIS, AND PLAUSIBLE EVIDENCE SUPPORTING THEIR RELATIONSHIPS WITH THE THREE DISEASES.

| GENES       | OBESITY                                                                 | CRC                                                                 | OSTEOPOROSIS                                                  |
|-------------|--------------------------------------------------------------------------|----------------------------------------------------------------------|---------------------------------------------------------------|
| PPP1R15A*   | In the bone morphogenetic protein (BMP) signaling pathway, which regulates appetite [34] | Mutations in the BMP pathway are related with colorectal carcinogenesis [35] | In the bone morphogenetic protein signaling pathway, which are associated with bone-related diseases, such as osteoporosis [36] |
| FOS         | diet-induced obesity is accompanied by alteration of FOS expression [37] | Proto-oncogene, in the KEGG pathway of colorectal cancer [38]         | Mice lacking c-fos develop severe osteopetrosis [39]          |
| FOSB        | positive association between maternal obesity [40]                      | Oncogene, regulators of cell proliferation, has a debatable impact on CRC patient survival [41] | Overexpression of FosB increases bone formation [42]         |
| HADHA*      | Associated with multiple fatty acid metabolism pathways [43]            | Unknown. Associated with breast cancer [44]                          | Unknown.                                                     |
| JUN         | The c-Jun NH2-terminal Kinase Promotes Insulin Resistance [45]           | Proto-oncogene, in the KEGG pathway of colorectal cancer [38]         | Associated with osteogenesis [46, 47]                        |
| NRIP1*      | Down-regulated in obese subjects, may suggest a compensatory mechanism to favor energy expenditure and reduce fat accumulation in obesity states [48] | Unknown. Involved in regulation of E2F1, an oncogene [49]          | Modulates transcriptional activity of the estrogen receptor. Interact with ESR1 and ESR2 in osteoporosis [50] |

* novel genes not involving insulin resistance pathways

C. Innovative Genes Shared Among Osteoporosis, Obesity and CRC Were Detected Using Gene Expression Meta-analysis

We downloaded five microarray series (GSE4017, GSE9348, GSE4183, GSE8671, GSE20916) for CRC, three (GSE48964, GSE29718, GSE55205) for obesity and three (GSE7429, GSE2208, GSE7158) for osteoporosis. Through meta-analysis, we obtained 9058 significant differentially expressed genes for CRC, 275 for obesity and 91 for osteoporosis. CRC and obesity shared a total of 192 genes. Among them, we found genes on insulin signaling pathways, such as PDK1, PRKAG2 and PDE3B, and adipocytokines, such as IL6 and IL8.

The three diseases osteoporosis, obesity and CRC shared six genes. Table II lists the genes and literature evidences, which support their relationships with each of the three diseases. Among them, FOS, JUN, and FOSB are oncogenes. FOS and JUN are known on the insulin signaling pathway. FOSB is on the AP1 pathway, which is associated with the proliferation of colon cancer cells [55]. Several studies suggested that overexpression of FOSB increases the responding of high fat reward while decreases energy expenditure and promotes adiposity [40, 56].

Interestingly, we found several genes not involving insulin signaling. Gene PPP1R15A is in the bone morphogenetic
protein signaling (BMP) pathway and its superfamily, the TGF beta signaling pathway. The mutation of BMP pathway has been found in patients with juvenile polyposis, which is rare syndrome with an increased risk for developing CRC [51, 52]. Mutations in TGF beta signaling also have been found susceptibility to CRC through genome-wide association studies [53]. A recent mouse experiment also showed that the BMP pathway regulates brown adipogenesis, energy expenditure and appetite, thus is highly associated with diet-induced obesity [54]. These evidences support our result. Further investigation is required to confirm and elucidate the role of the BMP pathway in the connection between obesity and CRC.

Gene NRIP1 regulates the estrogen receptor. Its interaction with sex hormone receptors plays a role in both obesity [48] and osteoporosis [50]. Its relationship with CRC is unclear yet, but studies suggested that estrogen may have protective effect on CRC [57]. Gene HADHA is on multiple pathways of fatty acid metabolism. But its role in CRC and osteoporosis is unknown yet.

To identify the common genes among obesity, CRC and osteoporosis, we currently analyzed the gene expression data, which can be noisy. While we found literature evidences to support the detected genes and their relationships with both obesity and CRC, these candidate genes need further investigations, for example, through mouse model experiments.

IV. DISCUSSIONS AND CONCLUSIONS

The genetic connection between CRC and obesity is multifactorial and inconclusive. In this study, we developed a comorbidity network analysis approach, which suggested that osteoporosis is important for the connection between obesity and CRC. We identified common genes among obesity, CRC and osteoporosis, and found these genes are associated with the regulation of sex hormone receptors and growth factors inducing bone formation. These genes are candidates in explaining the genetic overlaps between obesity and CRC.

Our comorbidity network may be not inclusive and biased toward the diseases whose drugs have high toxicity. The FDA adverse event reporting system collects data from medical product manufacturers, health professionals, and the public. The diseases without drug treatments are not included in the data, and the disease comorbidity relationships were often under-estimated in practice based on these data. In this study, we developed a network analysis approach to compensate the bias of the comorbidity data. In the future, including more complete patient disease data may facilitate the detection of new interesting comorbidities other than osteoporosis for obesity and CRC.

In addition, we currently detect comorbidities based on disease co-occurrence. The co-occurrence patterns may indicate the increase of the risk between two diseases in a mutual way. Incorporating more comprehensive patient-level data, such as time series data, may help refine the disease relationships and control confounding factors.

ACKNOWLEDGMENT

Our research was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number DP2HD084068, the training grant in computational genomic epidemiology of cancer (CoGEC) (R25 CA094186-06)

REFERENCES

[1] M. Oti, M. A. Huynen and H. G. Brunner. Phenome connections, Trends Genet, 24(3), 103–106.
[2] Blair, D. R., Lyttle, C. S., Mortensen, J. M., et al. (2013). A nondegenerate code of deleterious variants in mendelian Locis contributes to complex disease risk. Cell, 155(1), 70–80.
[3] Avery, C. L., He, Q., North, K. E., et al. (2011). A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. PLoS genetics, 7(10), e1002322.
[4] Calle, E. E., Rodriguez, C., Walker-Thurmond, K., and Thun, M. J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. New England Journal of Medicine, 348(17), 1629-1638.
[5] Bardou, M., Barkun, A. N., and Martel, M. (2013). Obesity and colorectal cancer. Gut, 62(6), 933-947.
[6] Khaodhiar, L., McCowen, K. C., and Blackburn, G. L. (1999). Obesity and its comorbid conditions. Clinical cornerstone, 2(3), 17-31.
[7] Pollak, M. (2008). Insulin and insulin-like growth factor signalling in neoplasia. Nature Reviews Cancer, 8(12), 915-928.
[8] LeRoth, D., and Roberts Jr, C. T. (2003). The insulin-like growth factor system and cancer. Cancer letters, 195(2), 127-137.
[9] Renehan, A. G., Zwahlen, M., Minder, C., O'Dwyer, S. T., Shalet, S. M., and Egger, M. (2004). Insulin-like growth factor (IGF)-1, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. The Lancet, 363(9418), 1346-1353.
[10] Dalamaga, M., Diakopoulos, K. N., and Mantzoros, C. S. (2012). The role of adiponectin in cancer: a review of current evidence. Endocrine reviews, 33(4), 547-594.
[11] An, W., Bai, Y., Deng, S. X., Gao, J., Ben, Q. W., Cai, Q. C., ... and Li, Z. S. (2012). Adiponectin levels in patients with colorectal cancer and adenoma: a meta-analysis. European Journal of Cancer Prevention, 21(2), 126-133.
[12] Wei, E. K., Giovannucci, E., Fuchs, C. S., Willett, W. C., and Mantzoros, C. S. (2005). Low plasma adiponectin levels and risk of colorectal cancer in men: a prospective study. Journal of the National Cancer Institute, 97(22), 1688-1694.
[13] Pärsttinen, R. P., Söderberg, S., Biessy, C., Ahrndt, B., Kaaks, G. H. R., and Olsson, T. (2003). Plasma leptin and colorectal cancer risk: a prospective study in Northern Sweden. Oncology reports, 10, 2015-2021.
[14] Tamakoshi, K., Toyoshima, H., Wakai, K., Kojima, M., Suzuki, K., Watanabe, Y., ... and Tamakoshi, A. (2005). Leptin is associated with an increased female colorectal cancer risk: a nested case-control study in Japan. Oncology, 68(4-6), 454-461.
[15] Zhu, Y., Michelle Luo, T., Jobin, C., and Young, H. A. (2011). Gut microbiota and probiotics in colon tumorigenesis. Cancer letters, 309(2), 119-127.
[16] Danese, E., Montagnana, M., Minicoczi, A. M., Bonafini, S., Ruzzene, O., Gelati, M., ... and Giudi, G. C. (2012). The role of resistin in colorectal cancer. Clinica Chimica Acta, 413(7), 760-764.
[17] Berster, J. M., and Göke, B. (2008). Type 2 diabetes mellitus as risk factor for colorectal cancer. Archives of physiology and biochemistry, 114(1), 84-98.
[18] Kominou, D., Ayonote, A., Richie, J. P., and Rigas, B. (2003). Insulin resistance and its contribution to colon carcinogenesis. Experimental Biology and Medicine, 228(4), 396-405.
[19] Karels, A. D. (2008). Metabolically healthy but obese individuals. Lancet, 372(9646), 1281-1283.
