

Interrogating Patient-level Genomics and Mouse Phenomics towards Understanding Cytokines in Colorectal Cancer Metastasis

Xiaoshu Cai, BS¹, Yang Chen, PhD², Chunlei Zheng, PhD², Rong Xu, PhD²

¹Department of Electrical Engineering and Computer Science, School of Engineering, Case Western Reserve University, Cleveland, Ohio, USA; ²Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

Abstract

Background: Colorectal cancer is the second leading cancer-related death worldwide and a majority of patients die from metastasis. Chronic intestinal inflammation plays an important role in tumor progression of colorectal cancer. However, few study works on systematically predicting colorectal cancer metastasis using inflammatory cytokine genes.

Results: We developed a supervised machine learning approach to predict colorectal cancer tumor progression using patient level genomic features. To better understand the role of cytokines, we integrated the metastatic-related genes from mouse phenotypic data. In addition, pathway analysis and network visualization were also applied to top significant genes ranked by feature weights of the final prediction model. The combined model of cytokines and mouse phenotypes achieved a predictive accuracy of 75.54%, higher than the model based on mouse phenotypes independently (70.42%, p -value<0.05). In additional, the combined model outperformed the model based on the existing metastatic-related epithelial-to-mesenchymal transition (EMT) genes (75.54% vs. 71.61%, p -value<0.05). We also observed that the most important cytokine gene features of the our model interact with the cancer driver genes and are highly associated with the colorectal cancer metastasis signaling pathway.

Conclusion: We developed a combined model using both cytokine and mouse phenotype information to predict colorectal cancer metastasis. The results suggested that the inflammatory cytokines increase the power of predicting metastasis. We also systematically demonstrated the critical role of cytokines in progression of colorectal tumor.

Keywords: cytokine, mouse phenotype, colorectal cancer metastasis, machine learning

Background

Metastasis is responsible for the majority of colorectal cancer mortality, and 90% patients with metastasis die within 5 years [1, 2]. However, the pathogenesis of metastasis remains poorly understood. Metastatic dissemination contains two phases: physical translocation and colonization at a secondary site [3, 4, 5]. The initial steps in the early phase are hard to detect and metastasis may lurk for years in most cases. This has become a complex situation for patients and can be hardly saved by surgeries [6]. Understanding the underlying mechanisms of metastatic process is an essential way to detect colorectal cancer metastasis and develop effective therapeutic strategies accordingly.

Inflammatory cytokines play a critical but disparate role in the colonization phase of colorectal cancer metastasis [7]. On the one hand, cytokines act as tumor antigen and curtail cancer progression [8]. On the other hand, cytokines conduct non-specific inflammatory activities can also promote cancer proliferation, invasion and metastasis [9]. Specifically, since the gut contains the highest density of microorganisms and has a well-developed immune system, chronic intestinal inflammation is a risk factor for colorectal cancer, showing a great potential for tumor progression [10, 11]. Recent studies identified mechanism of several critical inflammatory cytokine genes in CRC development [12-15], and emphasized the role of functional groupings of cytokines within complex networks [16]. However, few study works on systematic approach for inflammatory cytokine genes to predict CRC metastasis. Here, we investigated the combinational role of cytokines through combining patient genomics in The Cancer Genome Atlas (TCGA) and mouse phenomics data from Mouse Genome Informatics (MGI).

TCGA is a massive, comprehensive database that provides genomic data for over 20 types of cancers and clinical data for a large group of patients [17]. Gene expression profiling is a major component of data collected by TCGA. The data in TCGA is large-scale and well-organized, overcoming the limited sample size and incomplete available data of other relevant studies. Recently, existing systematic approaches have used the gene expression data and found high correlation between tumor prognostic and the epithelial-to-mesenchymal transition (EMT) genes [18-21]. The weak

signal in gene expression of cytokines make it challenging to predict CRC metastasis using cytokines independently. Generally, the pro-tumorigenic and anti-tumorigenic effects of cytokines are influenced by crosstalk with interacted genes in the complex cytokine milieu [7]. To help understand the role of cytokines, we addressed the problem by integrating external data. Specifically, the mouse model provides insights of gene functions through accessing to conserved processes such as metabolism, which is not available in human. Mouse Genome Informatics (MGI) [22], a recent widely-used mouse phenomics database, provides mouse phenotypic descriptions and ontology after systematic gene knockouts. In the past, these gene-phenotype associations in mice have been used for discovering new disease genes [23] and new drug effects [24]. In our previous study, we also leveraged both disease genetics and mouse phenotypes in MGI for drug repositioning [25-27]. In this study, combining mouse model phenotypic data and patient-level genomic data can facilitate detection of predictive gene panel and identify the power of cytokines.

In our study, we leveraged patient level data from TCGA and mouse phenotype level data from MGI to understand how cytokines mediated in colorectal cancer metastasis. We explored the combinational role of inflammatory cytokine genes through supervised machine learning. We integrated cytokines with the genes involving cancer metastasis and immune response in mice and built prediction models. The improved predictive power of the combination, compared with metastatic-related immune genes from MGI independently, shows a potential power of inflammatory cytokines in predicting CRC metastasis. In addition, our classification model outperformed the model of existing metastatic-related EMT genes. To further investigate the significance of cytokines in predicting CRC metastasis, network visualization and pathway analysis were also applied to top significant features in classification model.

Data and methods

Our experiments comprised of two steps: (1) combining cytokines and the mouse phenotype-gene association data to build classification models to predict colorectal cancer metastasis status; (2) understanding cytokines through visualizing top features of the final prediction model in interaction network and analyzing pathways of top features.

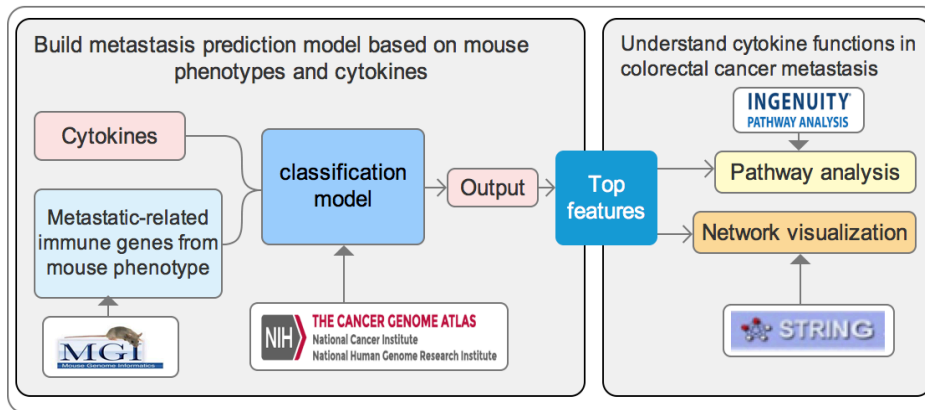


Figure 1. The overall workflow of our study

A. Extract public available colorectal cancer data (TCGA project)

The Cancer Genome Atlas (TCGA) provides high-throughput genome analysis for cancers. We exploited gene expression level data, as well as clinical meta-data from TCGA. Gene expression profiles generated using RNA-Seq of colorectal cancer were used, downloaded from TCGA [28] on the version of November 16th 2015. RNA-Seq expression dataset contains 328 patient samples on platform IlluminaHiSeq. Among them, with the knowledge of clinical features, we obtained 190 metastatic patients and 39 non-metastatic patients.

B. Build metastasis prediction model based on mouse phenotype data and cytokines

We collected available cytokine genes from The Immunology Database and Analysis Portal (ImmPort) from Cytokine Registry File on the version of November 2015. ImmPort [29] is an archival repository for clinical and molecular data with approximately 100 datasets now publicly available. Immune-related genes were downloaded from resource of Cytokine Registry, which is a master list of cytokines and receptors. For our purpose, we extracted all cytokine genes (without receptors, 145 in total).

To strengthen the predictive power of a gene panel related to cytokine that can classify colorectal cancer metastasis status, we integrated data from external data source Mouse Genome Informatics (MGI) [30] (obtained on Apr. 13th, 2016). MGI is a database that provides genetic, genomic and biological data from laboratory mouse. We used phenotype level data to find the genes with phenotype of metastasis-related in mouse model after being mutated. From literatures, in the physical phase of metastasis, process of angiogenesis provides a blood supply that support the metabolic needs [4, 6]. As a result, under the mammalian ontology, we considered the genes under categories: altered metastatic potential and abnormal angiogenesis, most related to metastasis in mouse model. We also downloaded mouse gene to human gene mapping file from MGI. After mapping, we obtained 122 altered metastatic potential genes and 244 abnormal angiogenesis genes.

In order to get immune response genes of all angiogenesis and metastasis genes from MGI, we used GO-term enrichment analysis on the Genome Ontology website [31]. GO-term enrichment analysis showed that “immune system process”, associated with 122 genes, and “immune response”, associated with 45 genes. In total, 146 genes, which is the union of “immune system process” and “immune response”, have the mouse phenotype of metastasis and are related to immune response. Furthermore, we combined cytokines with their interacted metastatic-related immune MGI genes as the candidate gene panel.

Existing systematic approaches have found high correlation between tumor progression and the epithelial-to-mesenchymal transition (EMT) genes. Accordingly, we obtained a total of 310 EMT genes from the study [19], which is a positive control in our evaluation. We compared our candidate gene panel with EMT genes in classifying metastasis.

We constructed the combination set of cytokines and metastatic-related immune genes from MGI and the model based on this set is our final prediction model. The other gene sets are evaluated for comparison. For all these five gene sets in Table 1, we used them as independent evaluation feature sets and performed classification using 10-fold cross-validation, respectively (we performed cross-validation due to the small sample size). The fold-change values of gene expression level data in TCGA were counted as input of classification. Since the RNA-Seq data for metastasis is unbalanced, we balanced the data using function ROSE in R. We performed the 10-fold cross-validation classifier for evaluation sets using SVM with polynomial kernel in WEKA [32]. For each set of genes, we randomly selected genes in the same size from more than 20000 genes, using the fold-change value in TCGA as input as well. We repeated the randomization process for 100 times and obtained statistical p-values using T-test.

Table 1. The number of genes in feature sets. Cytokines are from Immport Cytokine Registry list. Metastatic-related immune genes are from MGI. EMT genes are from a state-of-art related study [19].

| Feature set | Cytokines | Metastatic-related genes in MGI | Metastatic-related immune genes in MGI | Cytokines and their interacted metastatic-related immune genes in MGI and | EMT |
|--------------------|-----------|---------------------------------|--|---|-----|
| Number of features | 119 | 261 | 146 | 201 | 301 |

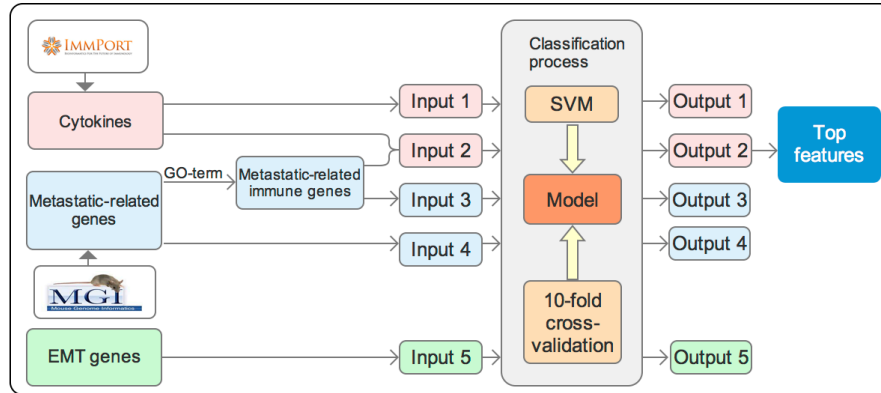


Figure 2. The workflow of classification analysis. The combination of cytokines and metastatic-related immune genes from MGI (input 2) is constructing the final prediction model. The other four models are evaluated as comparisons.

C. Understand cytokine functions in colorectal cancer metastasis

a. Network visualization

Network analysis provide a powerful tool to understand disease genetic mechanism [33-38]. To further discover the mechanism of cytokine genes in CRC metastasis, we explored the relationship between top 50 features selected by our final classification model and colorectal cancer driver genes through gene interaction network. We ranked the features by weight in our model, extracted the top 50 gene features, and mapped them into the protein interaction network. Then the full protein interaction data was downloaded from STRING [39] with an association score for each protein pair. In order to map the ENSP protein id to gene symbol, we obtained the full gene interaction network through mapping ENSP Id-Entrez Id file downloaded from STRING and Entrez Id-gene symbol file from customized download HUGO Gene Nomenclature (HGNC) [40]. Specifically, gene-gene interactions with low weight (score less than 500) were pruned to ensure a strong interaction sub-network. Furthermore, 13 colorectal cancer driver genes were curated from The National Cancer Institute (Gene Associated with a High Susceptibility of Colorectal Cancer, Table 2). As a result, a sub-network based on top features of the prediction model and colorectal cancer driver genes was constructed.

b. Network analysis and pathway analysis

Since the sub-network based on STRING is undirected, we integrated the information of regulation between genes to better understand the mechanism. First, we used personalized random walk to find genes that are directly interacted with top features. As a result, top features and their interacted genes were output after two steps of random walk on the whole gene interaction network. Second, we analyzed the regulation between top features and their interacted genes in network through Ingenuity Pathway Analysis. Ingenuity Pathway Analysis (IPA, <http://www.ingenuity.com>) is a web-based software that provides analytic tools such as ‘Canonical Pathway Analysis’ and ‘Upstream Regulator Analysis’, build on comprehensive ‘omics experiments. The set of top features and interacted genes was uploaded as input of IPA. The upstream regulator analysis reported the up-regulators and the genes it regulates among the input gene set. At last, we obtained a sub-network of top features of the combined model and cancer driver genes with the direction of regulation based on upstream analysis.

To analyze the pathway of top features selected from the combined model of cytokines and metastatic-related immune genes from MGI, we used the tool of ‘Canonical Pathway Analysis’ in IPA. The set of top features are input of IPA. The enrichment score and statistical significance p-value of involved pathways were calculated in Canonical Pathway Analysis. In order to have a better visualization of relevant pathways, we plotted top 20 pathways ranked by p-value.

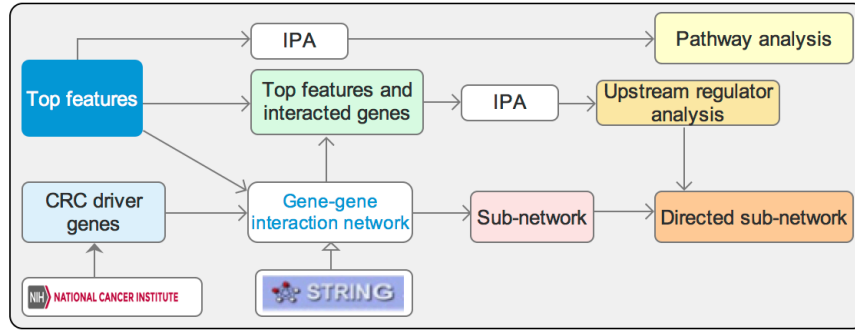


Figure 3. The workflow of network visualization and pathway analysis.

Results

A. A combined model of cytokine genes and mouse phenotypes is predictive of metastasis status in CRC patient

We evaluated the prediction performance of models based on five gene sets in Table 1. The results suggested that the model combining cytokines and metastatic-related immune genes achieved the highest predictive power among all. From Table 2, the model based on cytokines alone achieved an accuracy of 68.55%, which is significantly higher than that of randomly selecting genes in the same size (66.54%, p -value $<e$ -9, the randomly selecting genes can lead to higher than 50% accuracy since the random model can also have statistical significance in predicting CRC metastasis). We then evaluated the prediction power of mouse genes involving cancer metastasis. First, we tested all genes associated with mammalian phenotype of altered metastatic and abnormal angiogenesis in Mouse Phenotype Informatics. The accuracy in cross-validation is 69.43%, which confirmed the significance of mouse model metastatic genes in human. Second, we extracted all the genes belonging to immune response from the set of metastatic-related genes, which led to an accuracy of 70.42%. The size of the extracted gene set is a half smaller, but the predictive power is even higher than the original metastatic-related genes. The result suggests that immune response genes is representative in predicting metastasis.

Finally, we constructed a gene panel that combines cytokines and metastatic-related immune genes from MGI. The model based on combined features achieved a classification accuracy of 75.54%, which is significantly higher than that based on cytokines (p -value <0.01) or metastatic-related immune genes in MGI (p -value <0.05) alone. This model also significantly outperformed the model based on randomly selecting genes (p -value $<e$ -24). Our experiments have shown an increased power of predicting CRC metastasis when taking inflammatory cytokines.

Table 2. The measurements of evaluation set of cytokines, metastatic-related genes in MGI, metastatic-related immune genes in MGI and the combination of cytokines and metastatic-related immune genes in MGI. Correctly classified accuracy, precision and recall are reported. The p -value is the statistical significance compared with 100 times random selected genes in the same size.

| Feature set | Accuracy | Precision | Recall | P-value (compared with random select genes) |
|---|---------------|--------------|--------------|---|
| Cytokine genes | 68.55% | 0.686 | 0.686 | e -9 |
| Metastatic-related genes in MGI | 69.43% | 0.697 | 0.694 | 0.045 |
| Metastatic-related immune genes in MGI | 70.42% | 0.708 | 0.707 | e -5 |
| Cytokines and metastatic-related immune genes in MGI | 75.54% | 0.757 | 0.755 | e-24 |

To better evaluate the performance of our final prediction model, we also examined the classification model based on gene set of epithelial-to-mesenchymal transition (EMT) [19], an existing gene panel that is systematically demonstrated highly-correlated with colorectal cancer tumor progression. The classification accuracy of EMT genes is 71.75% and performed significantly better than random selecting genes (68.02%, p -value $<e$ -5). Table 3 shows that the combined model achieved a significantly higher performance in classifying patients with metastasis compared

with model of EMT genes (p -value <0.05). The result further demonstrated the significance of the combined features in our model. In summary, the result shows that cytokines play a significant role in predicting CRC tumor progression.

Table 3. The measurements of classification on model of existing metastatic-related EMT genes and our combined model. Correctly classified accuracy, precision and recall are reported. The p -value is the statistical significance compared with 100 times random selected genes in the same size.

| Feature set | Accuracy | Precision | Recall | P-value (compared with random select genes) |
|--|----------|-----------|--------|---|
| EMT genes | 71.61% | 0.716 | 0.716 | e-5 |
| Cytokines and metastatic-related immune genes in MGI | 75.54% | 0.757 | 0.755 | e-24 |

B. Cytokines gene panel offers insights into metastasis mechanisms

From the results above, we observed that the combined model has a higher potential to predict colorectal cancer metastasis. To understand the function of cytokines in this combined model, we visualized the interactions between the top genes of the model and colorectal cancer driver genes through network. In Figure 4, we assigned different color to three kinds of genes. Cytokine genes are represented in blue. Immune genes from MGI are colored in green. Red elements are colorectal cancer driver genes.

Figure 4 shows that significant cytokine genes including IL6, TGFB2, IFNG and CSF2, directly interact with metastatic-related immune genes and CRC driver genes. Specifically, one of the well-known colorectal cancer driver gene TP53 is mediated through the IL6, which is a classical player in proliferation, migration and angiogenesis [12, 13, 41, 42]. For another classical mediator in colorectal cancer TGFB2, it is regulated by cancer driver gene SMAD4 and also an up-stream regulator of SERPINE1, which is a potential epigenetic biomarker of colorectal cancer tumor progression [14, 15, 43, 44]. IL11, a new player in the colorectal cancer cytokine milieu, is regulated by IL6. IL11 have a strong correlation with STAT3 signaling and can be targeted therapeutically [45-48]. Granulocyte-macrophage colony stimulating factor (GM-CSF, gene symbol as CSF2 in Figure 4), which is used as an adjuvant to potentiate antitumor immunity in colorectal cancer [49, 50], is also mediated through IL6.

As for immune genes, IFNG, TGFB2, CSF2 and TP53 are up-regulators of FOS. In addition, ICAM1 is regulated by TP53, CSF2, IL13, IL2 and IL6. In all, the figure systematically shows the mechanism and regulation between cytokines, immune genes and cancer driver genes.

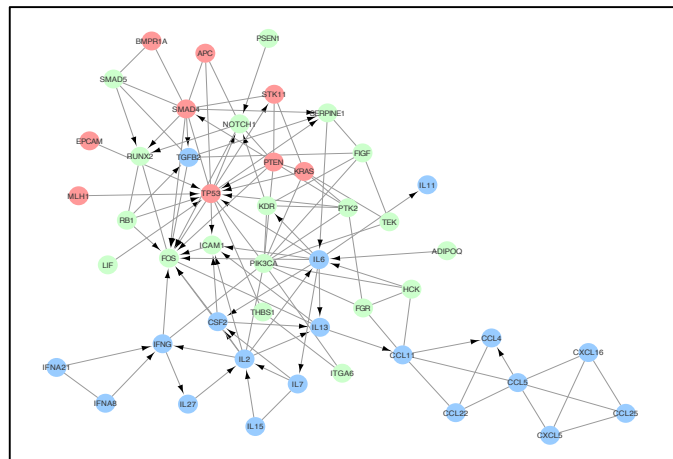


Figure 4. Sub-network of top 50 significant features and colorectal cancer driver genes. The arrow represents the direction of regulation. For example, SMAD4 is upstream regulator of TGFB2. The edge (without any arrow pointed) indicates the connection of a pair of genes in interaction network. Cytokine genes in blue; metastatic-related immune genes from mouse phenotype in green; colorectal cancer driver genes in red.

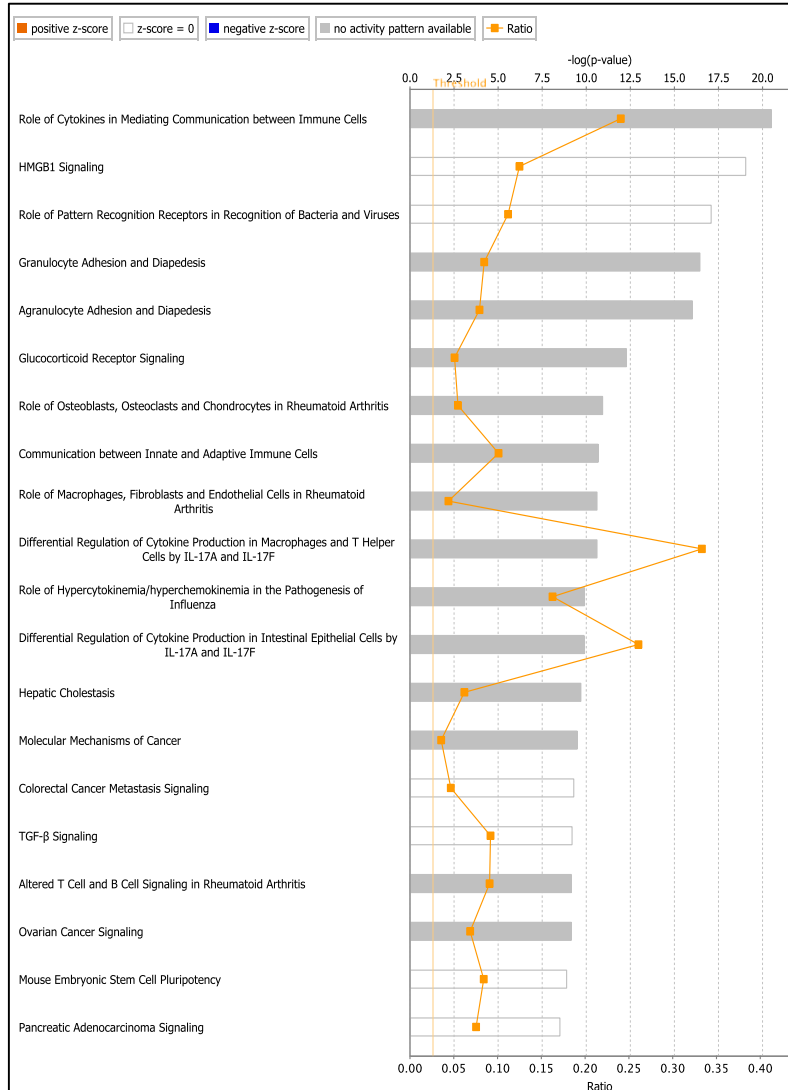


Figure 5. Ingenuity canonical pathway enrichment analysis of top significant features after classification. Top 20 canonical pathways ranked by $-\log(p\text{-value})$.

Additionally, we analyzed the pathway of top 50 features selected from the combined model through IPA. Top 20 canonical pathways are reported. Intriguingly, for example, TGF- β signaling pathway, ranked in top 20, plays a central role in predisposition and progression in colorectal cancer [51]. Most importantly, within top 20, colorectal cancer metastasis signaling pathway confirmed the combined model is closely correlated with metastasis of colorectal tumor from the aspect of mechanism.

Discussion

In our study, we explored the systematical mechanism and significance of inflammatory cytokine genes through the aspect of network visualization after classification on metastasis and non-metastasis. We still have space to improve our work in the following parts. First, we can integrate more kinds of data in the basis of classification except for the genes extracted from MGI. We used mouse genes with phenotype of angiogenesis and metastasis and mapped to human genes in this study. Besides, Cancer stem cells (CSCs) are thought to drive uncontrolled tumor growth and traits such as motility, invasiveness, and self-renewal, which are central to malignancy, may in fact be the reflection of the actions of the elusive CSC. These genes can all be integrated to our classification.

In addition, we currently considered all genes as independent in our classification model, while they are not in reality. Here, we investigated the role of cytokines in metastasis mechanism as a first step. In our future work, we will integrate the associations between genes and cover the discovery of genomic features of colorectal cancer metastasis to specific

organ. Through this way, more specific features will be observed and we can build a more comprehensive prediction model for colorectal cancer metastasis.

Furthermore, the selection of pivotal genes from classification can be used for drug repositioning for colorectal cancer in the future work. Based on the pipeline of computational method in discovering drugs using Library of Integrated Cellular Signatures (LINCS) data we explored in previous study [52], potential novel drugs for inhibition of growth of colorectal cancer metastasis can be found.

Conclusions

We have leveraged TCGA genomic data and mouse phenotypes in MGI to build a prediction model for colorectal cancer metastasis. We demonstrated the significance of inflammatory cytokine genes in classifying colorectal cancer metastasis by obtaining a better accuracy comparing with models based on other genetic features, including previously found metastasis-related gene panels. In addition, network and pathway analysis shows that significant cytokine genes in our prediction model are highly associated with cancer metastasis mechanisms. The results confirmed the critical role of cytokines mediated in CRC metastasis.

Abbreviations

CRC, colorectal cancer; MGI, mouse genome informatics; TCGA, the cancer genome atlas; EMT, epithelial-to-mesenchymal transition; CSC, cancer stem cells; LINCS, library of integrated cellular signatures.

Declarations

Funding

The project was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under the NIH Director's New Innovator Award number DP2HD084068.

Author's contributions

RX conceived the study. XC and YC designed the experiments. XC performed experiments, data analysis, and evaluation, and wrote the manuscript. All authors have participated in the study discussion and final manuscript review.

References

1. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M. J. Thun, Cancer statistics, 2008. *CA Cancer J. Clin.* 58, 71–96 (2008).
2. Jemal, Ahmedin, et al. "Global cancer statistics." *CA: a cancer journal for clinicians* 61.2 (2011): 69-90.
3. Chaffer, Christine L., and Robert A. Weinberg. "A perspective on cancer cell metastasis." *Science* 331.6024 (2011).
4. Folkman, Judah. "The role of angiogenesis in tumor growth." *Seminars in cancer biology*. Vol. 3. No. 2. 1992.
5. Chambers, Ann F., et al. "Clinical targets for anti-metastasis therapy." *Advances in cancer research* 79 (2000).
6. Chambers, Ann F., Alan C. Groom, and Ian C. MacDonald. "Metastasis: dissemination and growth of cancer cells in metastatic sites." *Nature Reviews Cancer* 2.8 (2002).
7. West, Nathan R., et al. "Emerging cytokine networks in colorectal cancer." *Nature Reviews Immunology* (2015).
8. Vinay, Dass S., et al. "Immune evasion in cancer: Mechanistic basis and therapeutic strategies." *Seminars in cancer biology*. Vol. 35. Academic Press, 2015.
9. Schreiber, Robert D., Lloyd J. Old, and Mark J. Smyth. "Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion." *Science* 331.6024 (2011): 1565-1570.
10. Jess, Tine, Christine Rungoe, and Laurent Peyrin-Biroulet. "Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies." *Clinical Gastroenterology and Hepatology* 10.6 (2012).
11. Lasry, Audrey, Adar Zinger, and Yinon Ben-Neriah. "Inflammatory networks underlying colorectal cancer." *Nature immunology* 17.3 (2016): 230-240.
12. Chung, Yuan-Chang, and Ya-Fen Chang. "Serum interleukin-6 levels reflect the disease status of colorectal cancer." *Journal of surgical oncology* 83.4 (2003).

13. Waldner, Maximilian J., Sebastian Foersch, and Markus F. Neurath. "Interleukin-6-a key regulator of colorectal cancer development." *Int J Biol Sci* 8.9 (2012).
14. Biswas, Swati, et al. "Transforming growth factor β receptor type II inactivation promotes the establishment and progression of colon cancer." *Cancer research* 64.14 (2004).
15. Becker, Christoph, et al. "TGF- β suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling." *Immunity* 21.4 (2004).
16. McInnes, Iain B., Christopher D. Buckley, and John D. Isaacs. "Cytokines in rheumatoid arthritis-shaping the immunological landscape." *Nature Reviews Rheumatology* (2015).
17. Cancer Genome Atlas Network. "Comprehensive molecular characterization of human colon and rectal cancer." *Nature* 487.7407 (2012): 330-337.
18. Shibayama, Masaki, et al. "Prediction of metastasis and recurrence in colorectal cancer based on gene expression analysis: ready for the clinic?" *Cancers* 3.3 (2011): 2858-2869.
19. Schell, Michael J., et al. "A composite gene expression signature optimizes prediction of colorectal cancer metastasis and outcome." *Clinical Cancer Research* (2015).
20. Loboda, Andre, et al. "EMT is the dominant program in human colon cancer." *BMC medical genomics* 4.1 (2011).
21. Tsai, Jeff H., and Jing Yang. "Epithelial–mesenchymal plasticity in carcinoma metastasis." *Genes & development* 27.20 (2013): 2192-2206.
22. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, The Mouse Genome Database Group. "The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease". *Nucleic Acids Res.* 2015, 28(43) (Database issue): D726-36.
23. Hoehndorf R, Schofield PN, Gkoutos GV. "PhenomeNET: a whole-phenome approach to disease gene discovery." *Nucleic Acids Research.* 2011;39(18): e119. doi:10.1093/nar/gkr538.
24. Kuhn, M., Banchaabouchi, M. A., Campillos, M., Jensen, L. J., Gross, C., Gavin, A.- C., and Bork, P. (2013). Systematic identification of proteins that elicits drug side effects. *Molecular Systems Biology*, 9, 663.
25. Chen Y, Cai X, Xu R (2015) Combing Human Disease Genetics and Mouse Model Phenotypes Towards Drug Repositioning for Parkinson's Disease. 2015 Annual American Medical Informatics Association Symposium, CA.
26. Chen Y, Gao Z, Wang B, Xu R. Towards precision medicine-based therapies for glioblastoma: interrogating human disease genomics and mouse phenotypes. *BMC genomics.* 2016 Aug 22;17(7):516.
27. Chen Y, Xu R. Drug repurposing for glioblastoma based on molecular subtypes. *Journal of Biomedical Informatics.* 2016 Dec 31;64:131-8.
28. TCGA, <https://tcga-data.nci.nih.gov/tcga>
29. Import, <http://import.niaid.nih.gov/>
30. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, The Mouse Genome Database Group. "The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease". *Nucleic Acids Res.* 2015, 28(43) (Database issue): D726-36.
31. GO-term, <http://geneontology.org>
32. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten. The WEKA data mining software: an update. *ACM SIGKDD Explor Newslett*, 11 (1) (2009).
33. Chen Y, Li L, Zhang GQ, Xu R. Phenome-driven disease genetics prediction toward drug discovery. *Bioinformatics.* 2015 Jun 15;31(12):i276-83.
34. Chen Y, Xu R. Network-based gene prediction for Plasmodium falciparum malaria towards genetics-based drug discovery. *BMC genomics.* 2015 Jun 11;16(7):1.
35. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics.* 2011 Jan 1;12(1):56-68.
36. Chen Y, Zhang X, Zhang GQ, Xu R. Comparative analysis of a novel disease phenotype network based on clinical manifestations. *Journal of biomedical informatics.* 2015 Feb 28;53:113-20.

37. Chen Y, Li L, Xu R. Disease Comorbidity Network Guides the Detection of Molecular Evidence for the Link Between Colorectal Cancer and Obesity. *AMIA Summits on Translational Science Proceedings*. 2015;2015:201.
38. Chen Y, Xu R. Mining cancer-specific disease comorbidities from a large observational health database. *Cancer informatics*. 2014 Jan 1(Suppl. 1):37.
39. Jensen, Lars J., et al. "STRING 8-a global view on proteins and their functional interactions in 630 organisms." *Nucleic acids research* 37. suppl 1 (2009): D412-D416.
40. HGNC, <http://www.genenames.org/cgi-bin/download>
41. Taniguchi, Koji, and Michael Karin. "IL-6 and related cytokines as the critical lynchpins between inflammation and cancer." *Seminars in immunology*. Vol. 26. No. 1. Academic Press, 2014.
42. Becker, C., et al. "IL-6 signaling promotes tumor growth in colorectal cancer." *Cell Cycle* 4.2 (2005): 220-223.
43. Mazzocchi, Gianluigi, et al. "ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer." *Journal of cancer research and clinical oncology* 138.3 (2012): 501-511.
44. Klein, R. Matthew, et al. "SERPINE1 expression discriminates site-specific metastasis in human melanoma." *Experimental dermatology* 21.7 (2012): 551-554.
45. Putoczki, Tracy L., et al. "Interleukin-11 is the dominant IL-6 family cytokine during gastrointestinal tumorigenesis and can be targeted therapeutically." *Cancer cell* 24.2 (2013): 257-271.
46. Ernst, Matthias, et al. "STAT3 and STAT1 mediate IL-11-dependent and inflammation-associated gastric tumorigenesis in gp130 receptor mutant mice." *The Journal of clinical investigation* 118.5 (2008): 1727-1738.
47. Yoshizaki, Ayumi, et al. "Expression of interleukin (IL)-11 and IL-11 receptor in human colorectal adenocarcinoma: IL-11 up-regulation of the invasive and proliferative activity of human colorectal carcinoma cells."
48. Putoczki, Tracy, and Matthias Ernst. "More than a sidekick: the IL-6 family cytokine IL-11 links inflammation to cancer." *Journal of leukocyte biology* 88.6 (2010): 1109-1117.
49. Nebiker, Christian A., et al. "GM-CSF production by tumor cells is associated with improved survival in colorectal cancer." *Clinical Cancer Research* 20.12 (2014): 3094-3106.
50. Wang, Yi, et al. "Tumor-derived GM-CSF promotes inflammatory colon carcinogenesis via stimulating epithelial release of VEGF." *Cancer research* 74.3 (2014): 716-726.
51. Xu, Yanfei, and Boris Pasche. "TGF- β signaling alterations and susceptibility to colorectal cancer." *Human molecular genetics* 16. R1 (2007): R14-R20.
52. Cai X, Chen Y, Gao Z, Xu R (2016) Explore Small Molecule-induced Genome-wide Transcriptional Profiles for Novel Inflammatory Bowel Disease Drug Discovery. *AMIA 2016 Joint Summits on Translational Science*.