

Exploration of *immunogene* in colon cancer recurrence

Qingrong Sun¹, Yao Chen¹, Shengyue Yuan¹ and Jun Liao^{1,2}

¹School of Science, China Pharmaceutical University, Nanjing, China

²Key Laboratory of Drug Quality Control and Pharmacovigilance, China Pharmaceutical University, Ministry of Education, Nanjing, China

Correspondence to: Jun Liao, **email:** liaojun@cpu.edu.cn

Keywords: *gene expression profiling; pathways; recurrence; somatic mutation; TCGA*

Received: April 21, 2017

Accepted: January 09, 2018

Published: January 12, 2018

Copyright: Sun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Colon adenocarcinoma is the third most common cancer with high risk of recurrence and deteriorative consequences. Given the importance of *immune genes* in tumor regulation and cancer immunotherapy, there is a need to comprehensively profile the immunoregulatory genes from multiple types of colon cancer patient genomic data for discovering important associations and potential therapeutic targets of colon cancer recurrence. We used publicly available colon tumor tissue genomic data from The Cancer Genome Atlas database and immune genes data from innateDB database in this study. We derived the immune genes profiles by exploring multiple genomic profiles (gene expression, clinical and somatic mutation) in colon cancer. Some of the synthetic lethal genes we identified, such as CASP14, MS4A6E, KIR2DL1, KIR3DL1, KIR2DL3, CCL1, IL36B, FOXO3, POU2F1, SMAD3, HOXA9, PACS1, PROM1, DIDO1, SRC, CBFA2T2, NCOA6, PGAM1 and PROC, have been suggested to be potential targets correlated with immune genes for colon cancer recurrence treatment. Moreover, TLR2 could be promisingly new early stage indicator for colon adenocarcinoma recurrence. This is a systematic study that combines three different types of genomic data to molecularly characterize colon cancer and aims to identify potential targets for colon adenocarcinoma therapy. Meanwhile, the integrative analysis of immune genes for colon cancer could assist in identifying potential new symbols for colon adenocarcinoma recurrence.

INTRODUCTION

Colon adenocarcinoma (COAD), as a kind of colorectal cancer, has traditionally been treated surgically [1]. However, many cases of colon cancer are systemic at the time of diagnosis, and apparently curative surgery is turned out to be at a late stage. But, tumor recurrence as a consequence of circulating tumor cells is unmanageable before the surgery [2–4]. It has been recognized that cancer is associated with the genetic, genomic and epigenetics changes [5, 6]. Meanwhile, a demonstrative influence of the host immune response on tumour invasion, recurrence and metastasis has come from analyses of the *in situ* immune components and how these components are organized within human tumours. Indeed, immune infiltrates are heterogeneous between tumour types, and are very diverse from patient to patient. All immune cell types may be found in a tumour, including macrophages,

dendritic cells, mast cells, natural killer (NK) cells, naive and memory lymphocytes, B cells and effector T cells (including various subsets of T cell: T helper cells, T helper 1 (T_H1) cells, T_H2 cells, T_H17 cells, regulatory T (T_{Reg}) cells, T follicular helper (T_{FH}) cells and cytotoxic T cells). These immune cells can be located in the core (the centre) of the tumour, in the invasive margin or in the adjacent tertiary lymphoid structures (TLS) [7]. In some cases, immune cells constitute an additional, prominent component of the host response to cancer, but their participation in tumour pathogenesis remains incompletely understood [8]. Therefore, finding genes correlated with immunogenes for COAD recurrence is becoming more and more important.

A large number of COAD genomic data are emerging and promoting colon cancer research. The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) gave comprehensive molecular portraits of human

colon cancer by integrating various types of “omics” data including genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA (miRNA) sequencing, and reverse-phase protein arrays. The related investigations have greatly advanced our understanding of colon cancer in molecular profiles [9], although translation of genomic findings into clinical applications remains challenging. The high-quality TCGA primary colon tumor samples and their comprehensive molecular profiles could be an invaluable source of information for molecular exploration of colon cancer and discovery of new treatment targets. Immune gene list from InateDB, as a manually-curated knowledgebase of the genes, proteins, and particularly, the interactions and signaling responses involved in mammalian innate immunity [10], include Immport [11], Immunogenetic Related Information Source (IRIS), Septic Shock Group, MAPK/NFKB Network, Calvano et al., Nature 2005 [12], and Immunome Database. Considering the dependability of all immune genes, we selected immune genes that are distributed in more than two above gene lists, and then incorporated the innate immunity genes that are not distributed in other gene lists.

Microsatellite instability (MSI) is caused by a defect in the mismatch repair (MMR) machinery, one of the main mechanisms responsible for recognizing and repairing errors in newly synthesized DNA. MSI results from biallelic inactivation of one of the MMR genes [13]. MSI, as one of immunogenic subtype in Colorectal cancer, has been explored in research paper [14, 15]. Mutation burden defined by the number of mutated genes in every cancer sample mirrors the degree of mutation [16]. The relationship of mutations and MSI or gene expression level and MSI need to be analyzed.

In this study, we explored immune genes correlated with COAD recurrence by survival analyses based on immunogene mutations profiles. We analyzed immune gene somatic mutation and gene expression data to identify potential SL genes for above immune genes, evaluated MSI status, mutation burden and compared the relation of MSI with somatic mutation and gene expression level. We also identified potential sign for recurrent COAD by immunogene mutations and clinical profiles and verified the results by PubMed references (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>).

RESULTS

By using Gene Set Enrichment Analysis (GSEA) software [17], we identified 10 canonical pathways significantly associated with this 1219-gene set that have significant differences between recurrent cancer and normal samples. We obtained 10 pathways correlated with 1219 Immune genes involved in COAD by a threshold of adjusted *P*-values (FDR) < 0.05 (Figure 1 and Supplementary Table 1), such as Genes involved

in Immune System [18], Matrisome and Matrisome Associated [19, 20], Innate Immune System, Complement and Coagulation Cascades, Adaptive Immune System, Pathways in cancer, Hemostasis, Hematopoietic cell lineage, Cytokine–cytokine receptor interaction [21].

We compared disease-free survival (DFS) between immune genes mutated and wildtype cancer samples, both of more than 3 samples in COAD. Finally, 69 Genes were found that all have a worse DFS prognosis, and then are bad for recurrent COAD patients to survive. We plot DFS curves for immune gene-mutated and gene-wildtype cancer samples with significant poor prognosis by a threshold of adjusted *P*-value (FDR) < 0.05 (Figure 2).

According to the MSI status, we divided all recurrent COAD samples and all non-recurrent cancer samples into three groups, including MSI group, Microsatellite stability (MSS) group, and non-available group. Then, we concluded that the COAD recurrence has no correlation with the stability of microsatellites.

DNA mutation types were classified by TCGA mutation dataset, which included Silent, Missense_Mutation, Frame_Shift_Del, Nonsense_Mutation, In_Frame_Ins, Splice_Site, Frame_Shift_Ins, In_Frame_Del (Figure 3 and Supplementary Table 2). We found that higher mutation burden of recurrent cancer samples concentrated in missense mutation except for silent mutation.

Based on the TCGA data, the stability of microsatellites has no significant correlation with cancer recurrence. Then, we concluded that immunogene cell adhesion molecule L1-like (CHL1, *P*-value = 0.019, odds ratio = 7.644) and insulin receptor substrate 1 (IRS1, *P*-value = 0.026, odds ratio = Inf (denominator is zero)) somatic mutation have correlation with the MSI and MSS. At the same time, the differences of expression levels of the 69 immunogenes in MSI cancer samples compared to those in MSS cancer samples were shown in Table 1 (Supplementary Table 3).

Comparison expression level of immune genes that are correlated with recurrent COAD in recurrent cancer and normal samples, which shows that 31 immune genes have significant difference between the two groups (FDR < 0.05, and fold change > 1.5). The FDR was estimated using the method of Benjami and Hochberg [22]. Heat map of 31 immune genes expression (Figure 4 and Supplementary Table 4) shows the expression trend in recurrent cancer and normal samples.

Figure 4. Differential gene expression levels of immune genes in recurrent cancer compared to those in normal samples. The columns represent recurrent cancer samples (63), normal samples (41) without cancer, and the rows represent immune genes. The red color indicates that a gene is more highly expressed, and the green color indicates that it isn't.

We verified the 31 immune genes correlated with COAD recurrence by using the PubMed database. All

Table 1: The differences of expression levels of gene-set from 69 immunogenes in MSI cancer samples compared to those in MSS cancer samples

Symbol	Description	P-value	Log2(Fold change ^a)
IL17C	interleukin	2.25E-11	3.168963
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	4.03E-06	0.868329
CD99L2	CD99 molecule-like 2	0.001053	-0.68528
PGM5	phosphoglucomutase 5	0.001941	-2.1586

^afold change = mean MSI cancer samples expression levels/mean MSS cancer samples expression levels.

research papers related to the 31 genes provided direct or indirect evidences suggesting that COAD recurrence is affected by immunogenes (Table 2). We found that the related experiments of the I^2 immune genes have been published in research papers.

We identified a set of candidate synthetic lethal (SL) genes [38] for 31 immune genes from above results. 19 SL genes for 6 immune genes have been found in mutation data (Table 3).

Moreover, we compared IC_{50} (drug concentration that reduces viability by 50%) values between immune genes that identified for 19 SL genes' higher-expression-level and lower-expression-level cancer cell lines for

each of the 265 compounds mentioned above. We found that GULP1 had significantly lower IC_{50} values in lower-expression-level cancer cell lines than in higher-expression-level cancer cell lines, and other 4 immune genes turned out to be the reverse (P -value < 0.05 and top 5 ascending order by FDR) (Table 4 and Supplementary Table 5). This indicates that immune genes expression levels of the cancer cell lines significantly influence the sensitivity of these cell-lines to drugs. Table 4 lists drug sensitivity differences in differential expression levels of immune genes obtained by the Cancer Cell Line Project.

We compared the immune gene mutation rates among different clinical phenotypes of cancer patients

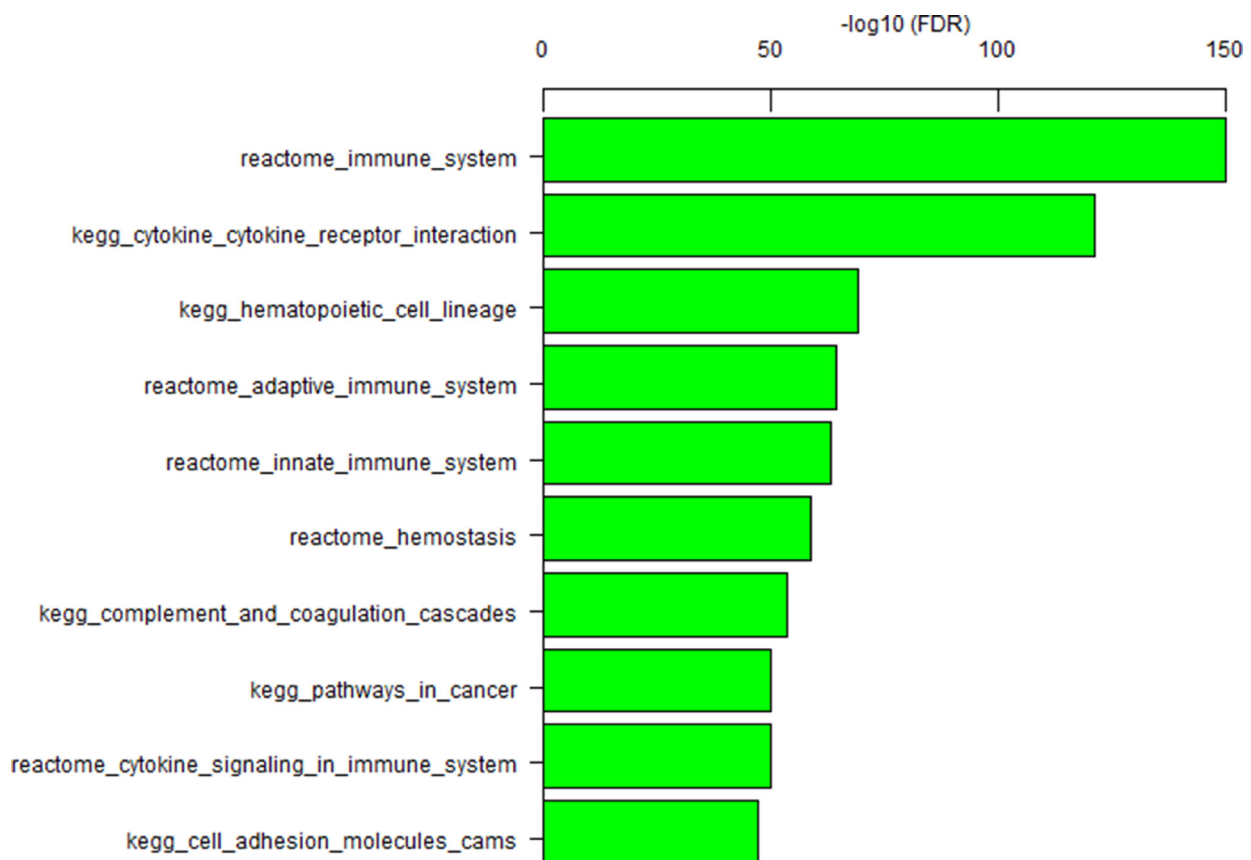


Figure 1: Important biological functions associated with 1219 immune genes.

Table 2: The evidences for verifying the relationship between immunogenes that are correlated with CAOD recurrence and have significant differences between in recurrent cancer samples and in normal samples with cancer recurrence

Symbol	Description	Reference
RIPK1	receptor (TNFRSF)-interacting serine-threonine kinase 1	[23–25]
CHL1	cell adhesion molecule L1-like	[26]
CEBPβ	CCAAT/enhancer binding protein (C/EBP), beta	[27]
CHD7	Chromodomain Helicase DNA Binding Protein 7	[28]
CYSLTR1	cysteinyl leukotriene receptor 1	[29]
CD96	CD96 Molecule	[30]
HLA-C	major histocompatibility complex, class I, C	[31]
IL17C	interleukin 17C	[32]
NLRC3	NLR family, CARD domain containing 3	[33, 34]
IL18R1	interleukin 18 receptor 1	[35]
GULP1	GULP, engulfment adaptor PTB domain containing 1	[36]
EOMES	eomesodermin	[37]

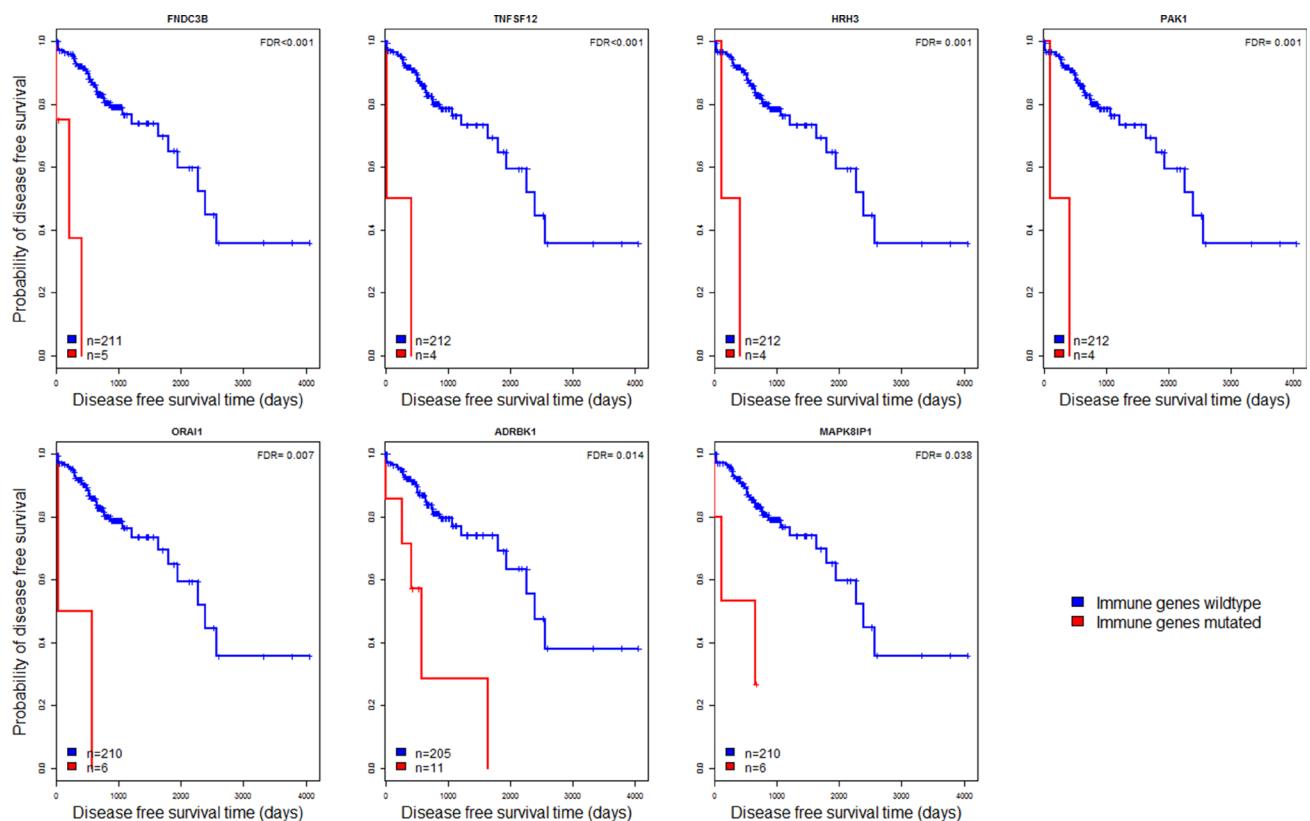


Figure 2: Kaplan–Meier survival curves, which show significant disease-free survival (DFS) time differences between immune gene-mutated and immune gene-wildtype cancer patients (log-rank test, FDR < 0.05).

Table 3: The 19 genes that are potentially synthetic lethal for immune genes that are correlated with COAD recurrence

Symbol ^b	Description	Symbol ^c	Pathway ^d	Drug ^e
CASP14	caspase 14	ATP5A1, IL18R1	NA	Apoptosis Activator 2, Boc-D-FMK, PAC-1, MDL 28170, PD 150606
MS4A6E	membrane spanning 4-domains A6E	ATP5A1	NA	NA
KIR2DL1	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 1	IL18R1, CD96	Immune System, Adaptive Immune System	NA
KIR3DL1	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1	IL18R1	Immune System	NA
KIR2DL3	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3	IL18R1	Immune System	NA
CCL1	chemokine (C-C motif) ligand 1	IL18R1	NA	NA
IL36B	interleukin 36, beta	GULP1	Immune System,	NA
FOXO3	forkhead box O3	LYL1, SPP1	Signaling by PDGF, Signal Transduction	NA
POU2F1	POU class 2 homeobox 1	LYL1	NA	NA
SMAD3	SMAD family member 3	LYL1	NA	NA
HOXA9	homeobox A9	LYL1	NA	NA
PACS1	phosphofurin acidic cluster sorting protein 1	LYL1	NA	Mannose 6-phosphate
PROM1	prominin 1	LYL1	NA	NA
DIDO1	death inducer-obliterator 1	SPP1	NA	NA
SRC	SRC proto-oncogene, non-receptor tyrosine kinase	SPP1	Signaling by PDGF, Signal Transduction, Membrane Trafficking Signaling by GPCR	Dasatinib, bosutinib, ponatinib, Nintedanib, Bevacizumab, Bosulif, Iclusig, Ofev, Sprycel, Adenosine triphosphate
CBFA2T2	core-binding factor, runt domain, alpha subunit 2; translocated to, 2	SPP1	NA	NA
NCOA6	nuclear receptor coactivator 6	SPP1	NA	Cholecalciferol
PGAM1	phosphoglycerate mutase 1 (brain)	SPP1	NA	Phosphoric acid, Water
PROC	protein C (inactivator of coagulation factors Va and VIIIa)	SPP1	Cell surface interactions at the vascular wall, Post-translational protein modification	Sodium Tetradecyl Sulfate, Menadione, Warfarin, Antihemophilic Factor (Recombinant), Drospirenone, Ethinyl Estradiol, Norelgestromin, Phenprocoumon, calcium

^bSL gene symbol for immune genes, ^cImmune genes symbol, ^dPathways to which the kinase gene is related, ^eDrugs that have been approved, ^fData on Pathways and Compounds from the GeneCards (www.genecards.org), KEGG (www.genome.jp/kegg/), REACTOME (www.reactome.org/)

using Fisher's Exact Test. Stage phenotype that represents tumor size and spread was divided into two classes: early stage (Stage I-II) vs. late stage (Stage III-IV) [50]. Only gene TLR2 mutation was correlated with stages, and its

mutation rate was higher in early stage than late stage subjects (unadjusted *P*-value = 0.041, odds ratio = Inf). Gene toll-like receptor 2 (TLR2) mutation rate is 3.7%, that is, 8 samples, which include Missense Mutation (3),

Table 4: The drug sensitivity differences in differential expression levels of immune genes obtained by the cancer cell line project

Symbol ^f	Description	Compound ^g	Sensitivity ^h
ATP5A1	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	AZD6482 [39, 40], AV-951 [41, 42], MG-132, Dasatinib [43], SB-505124	A > B
CD96	CD96 molecule	Temsirolimus, AZD6482 [39, 40]	A > B
IL18R1	interleukin 18 receptor 1	PLX4720-rescreen [44], BX-795 [45], NU-7441, Bosutinib [46], MLN4924 [47]	A > B
GULP1	GULP, engulfment adaptor PTB domain containing 1	X5-Fluorouracil, KIN001-135, SNX-2112 [48]	A < B
LYL1	lymphoblastic leukemia associated hematopoiesis regulator 1	GSK1070916 [49]	A < B
SPP1	secreted phosphoprotein 1	AZD6482 [39, 40], AV-951 [41, 42], MG-132, Dasatinib [43], SB-505124	A > B

^fSymbols of immune genes that have SL genes., ^gCompounds from the Cancer Cell Line Project (www.cancerrxgene.org/) that rank Top 5 in ascending order by FDR and *P*-value < 0.05., ^hA: higher-expression-level, B: lower-expression-level.

Frame Shift Insert Mutation (2), Nonsense Mutation (1) and Silent Mutation (2).

DISCUSSION

In this study, we performed extensive analyses of immune gene mutation, gene expression, and clinical data from COAD datasets in TCGA. We computed the somatic mutation burden, MSI status and correlation with each other of 69 immunogenes, verified the immunogenes correlated with COAD recurrence by literature, identified potential druggable SL partners of immune genes that are correlated with COAD recrudescence, analyzed correlation between early and late stages.

Druggable SL gene partners that were identified for immune genes that are correlated with COAD recidivation may yield insights into the personalized treatment of patients with immune gene-mutated cancers, since no druggable immune gene mutants. An example of successful application of the synthetic lethality approach is the targeting of cancers with dysfunction of the breast-cancer susceptibility genes 1 and 2 (*BRC1* and *BRC2*) by poly(adenosine diphosphate *ADP*-ribose) polymerase (*PARP*) inhibitors [51]. In the present study, we identified potential immune genes SL partners based on the assumption that mutation of SL gene has more effects on the expression of immune genes in *SL gene*-mutated cancers than those in both *SL gene*-wildtype cancers and normal tissue. Moreover, we validated the correlation between drug sensitivity and gene expression by exploring the pharmacogenomic data from the Cancer Cell Line Project. We identified a threshold of *P*-value < 0.05 in the result of different sensitivity based on the pharmacogenomic data. Drugs had significant differences between higher-expression-level and lower-expression-level genes. ATP5A1, IL18R1, and SPP1 all have higher

sensitivity in higher-expression-level than lower-expression-level. GULP1 and LYL1 all have lower sensitivity in higher-expression-level than lower-expression-level. CGP-082996 [52], as inhibitor of CDK4, is the only one drug that has lower sensitivity in higher-expression-level than in lower-expression-level based on different drug sensitivity of ATP5A1. We identified that CDK4 expression significantly correlates with ATP5A1 expression in COAD (Pearson product-moment correlation, *P*-value < 0.05). The result is based on COAD dataset from TCGA indicating that CDK4 expression positively correlates with ATP5A1 expression (*P*-value = 0.004, Correlation coefficient = 0.221), and then explains the reason of lower drug sensitivity in ATP5A1 higher-expression-level.

TLR2 plays an important role in Lewis lung carcinoma metastatic growth [53]. Recurrence pattern of COAD includes metastatic recurrence [54] and disseminated recurrence [55]. TLR2 is also required for rapid inflammasome activation [56, 57] and is identified as an indication gene correlated with cancer stages and cancer recurrence by this study. TLR2 interacts with a number of gene products (proteins) (Figure 5, generated by the BioGRID [58]). For example, autosomal recessive deficiencies of IRAK1 and MYD88 impair Toll-like receptor (TLR) to recurrent life-threatening bacterial diseases [59]. TIRAP is dispensable in TLR2 signalling at high ligand concentrations in macrophages and dendritic cells, with MyD88 probably coupling to the TLR2 receptor complex at sufficient levels to allow activation but having an inhibitory role in the signalling of TLR3 to JNK [60, 61]. Although these results need to be validated through experimental investigation, they represent a promising direction for future studies.

Overall, we found 31 immune genes related to colon cancer recurrence. In particular, 12 of the 31 identified genes have been reported in the literatures to

be related to colon cancer recurrence [23–37], which can be regarded as validation to our prediction results for these 12 genes. For the other 19 genes identified by us, we cannot exclude the possibility that some of them are due to false positive prediction. Therefore, there is a need for further experimental studies and external database of our predicted genes to validate our method and the results.

Our study was based on the data from the TCGA database. It is desirable to test and apply our method on the data from other sources. We therefore searched for the relevant data from other databases including Gene Expression Omnibus (GEO) [62], Ensembl [63], Hugo Gene Nomenclature Committee (HGNC) [64], ArrayExpress [65], and Catalogue of Somatic Mutations in Cancer (COSMIC) [66]. These databases provide comprehensive genome data for colon cancer, such as gene expression dataset, somatic mutation dataset and gene location dataset even gene Methylation dataset. However, none of them provide the information of the clinical prognosis tracking, which is needed by our study. Hence, we were unable to test and apply our method on the data from these databases. It is expected that more comprehensive information such as clinical prognosis tracking will be made available in more databases, which will better serve the need of various research and discovery efforts.

MATERIALS AND METHODS

Materials

We downloaded the colon carcinoma gene expression (microarray), gene somatic mutation data and clinical dataset from the TCGA data portal ([https://](https://portal.gdc.cancer.gov/)

portal.gdc.cancer.gov/). For TCGA samples, somatic mutations were revealed from exome sequencing of matched tumor and normal tissue genome pairs. In the gene expression data, a total of 287 colon cancer samples and 41 normal samples were found. Considering that the gene expression activity is our primary concern, and for statistical consistency, we analyzed the same 287 colon cancer samples in the other 2 data types. There are 79 and 4 samples missing in gene somatic mutation and clinical data, respectively. Thus, we analyzed 287 colon cancer samples in the gene expression, 208 colon cancer samples in the gene mutation data and 283 colon cancer in the colon cancer clinical data. We used clinical data for survival analyses and indication of tumor status from FireBrowse (<http://gdac.broadinstitute.org/>). According to the dataset mentioned above, 63 recurrent tumor samples and 177 non-recurrent tumor samples have been divided into two cancer groups. Considering that the TCGA dataset activity is our primary concern, and for statistical consistency, we analyzed 2546 immune genes contained in COAD from InnateDB (<http://www.innatedb.ca/redirect.do?go=resourcesGeneLists>) [10]. We obtained pharmacogenomic data from the Cancer Cell Line Project (<http://www.cancerrxgene.org/>) covering 265 screened compounds and their targets, and including cancer cell lines' drug response, drug sensitivity, and gene expression data. Ethical approval was avoided since we used only publicly available data and materials in this study.

Comparison of immune genes expression levels

We normalized TCGA RNA-Seq gene expression data by base-2 log transformation. We identified differentially expressed genes between two classes of

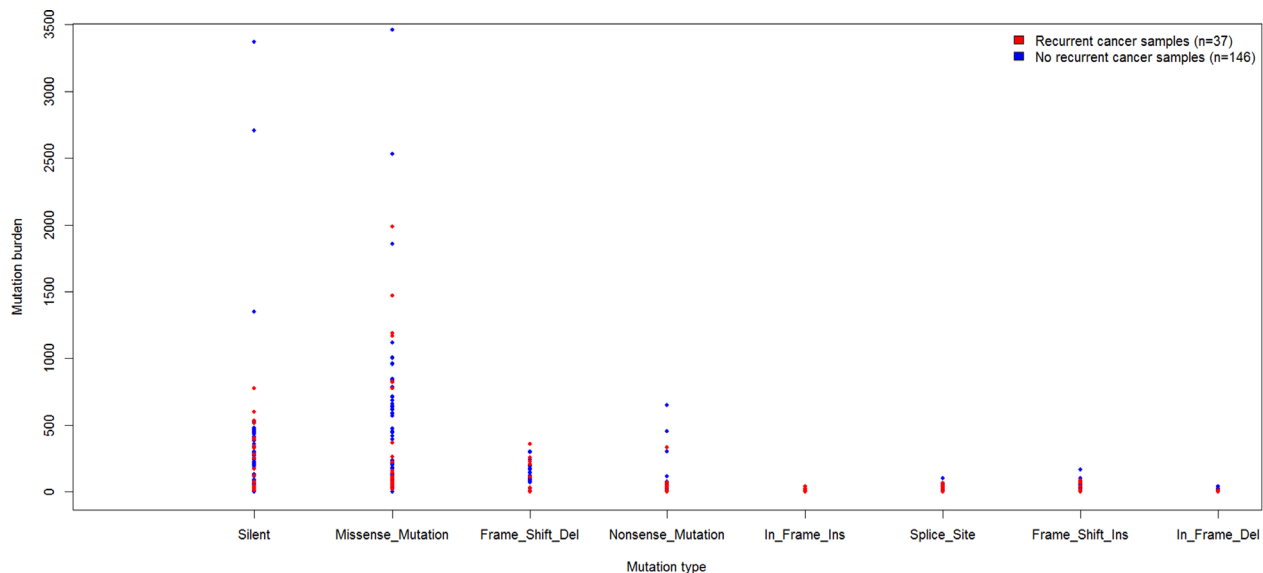


Figure 3: Scatter graph for somatic mutation burden of cancer samples (recurrent cancer and non-recurrent cancer). Every dot represents a sample, wherein, the red dots are recurrent cancer samples and blue dots are non-recurrent cancer samples. The vertical axis shows the number of mutation genes in every cancer sample.

samples using Student's *t* test. To adapt to multiple tests, we calculated adjusted *P*-values (FDR) for *t* test *P*-values. We used the threshold of FDR < 0.05 and mean gene-expression fold-change > 1.5 to identify the differentially expressed genes.

Comparison of the immune gene mutation rates and expression levels

We compared the immune gene mutation rates among different clinical phenotypes of cancer patients using Fisher's Exact Test. Tumor stage phenotype was

divided into two classes: early stage (Stage I-II) vs. late stage (Stage III-IV). A threshold of *P*-value < 0.05 was used to evaluate the correlation in mutation rates between the two classes of phenotypes.

Gene-set enrichment analysis

We performed pathway analysis of the gene sets using KEGG(www.genome.jp/kegg/), REACTOME (www.reactome.org/) and the GSEA tool (<http://software.broadinstitute.org/gsea/msigdb/>). We carried out network analysis of gene sets interacting with a

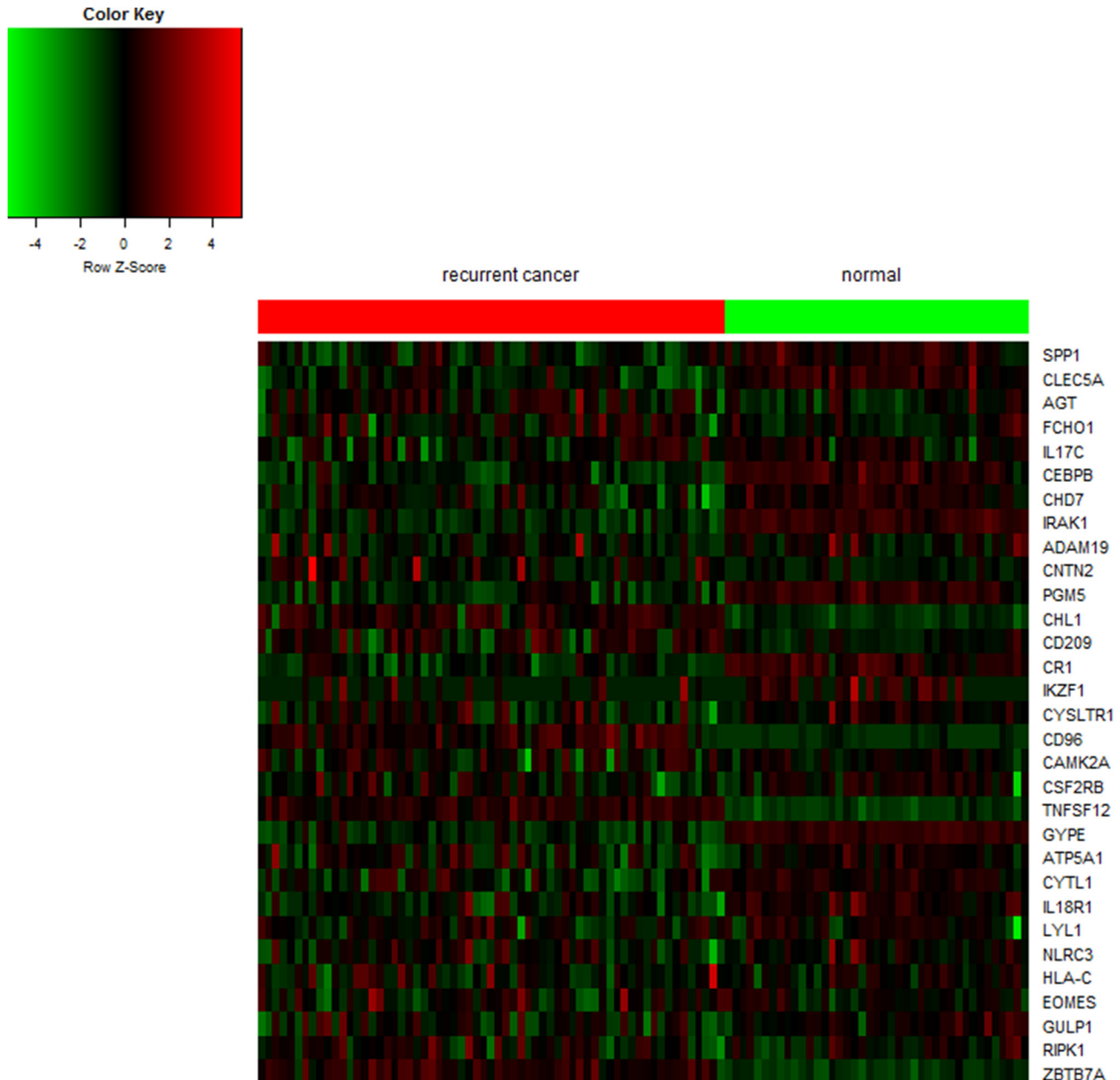


Figure 4: Differential expression levels of immune genes in recurrent cancer compared to those in normal samples. The columns represent recurrent cancer samples (63), normal samples (41) without cancer, and the rows represent immune genes. The red color indicates that a gene is more highly expressed, and the green color indicates that it isn't.

number of gene products (proteins) generated by the BioGRID [58].

Survival analyses

We performed survival analyses of TCGA patients based on COAD mutation data. Kaplan–Meier survival curves were used to show the survival (disease free survival) differences between gene-mutated cancer patients and gene-wildtype cancer patients. We used the log-rank test to calculate the significance of survival-time differences between the two classes of patients with a threshold of P -value < 0.05 .

Microsatellite instability status in all cancer samples

We divided the MSI into three groups, including MSI, MSS and non-available, and analyzed the correlation between stability of microsatellites and somatic mutation by using Fisher’s Exact Test. Based on the cancer recurrence immunogenes, we compared the correlation of 69 genes between gene somatic mutation and MSI by using Fisher’s

Exact Test (P -value < 0.05) in COAD. At the same time, we analyzed the differential expression levels in MSI cancer samples compared to those in MSS cancer samples by Student’s t test with a threshold of FDR < 0.05 and a fold change > 1.5 .

Somatic mutation burden of all gene-mutated samples

We integrated the clinical data and somatic mutation data. Then, we showed mutation burden and mutation types of all recurrent cancer samples.

Verification of genes by the research based on the PubMed

We obtained related research papers of experimental data from PubMed searches, and integrated all genes information to verify the 31 immune genes correlated with COAD recurrence.

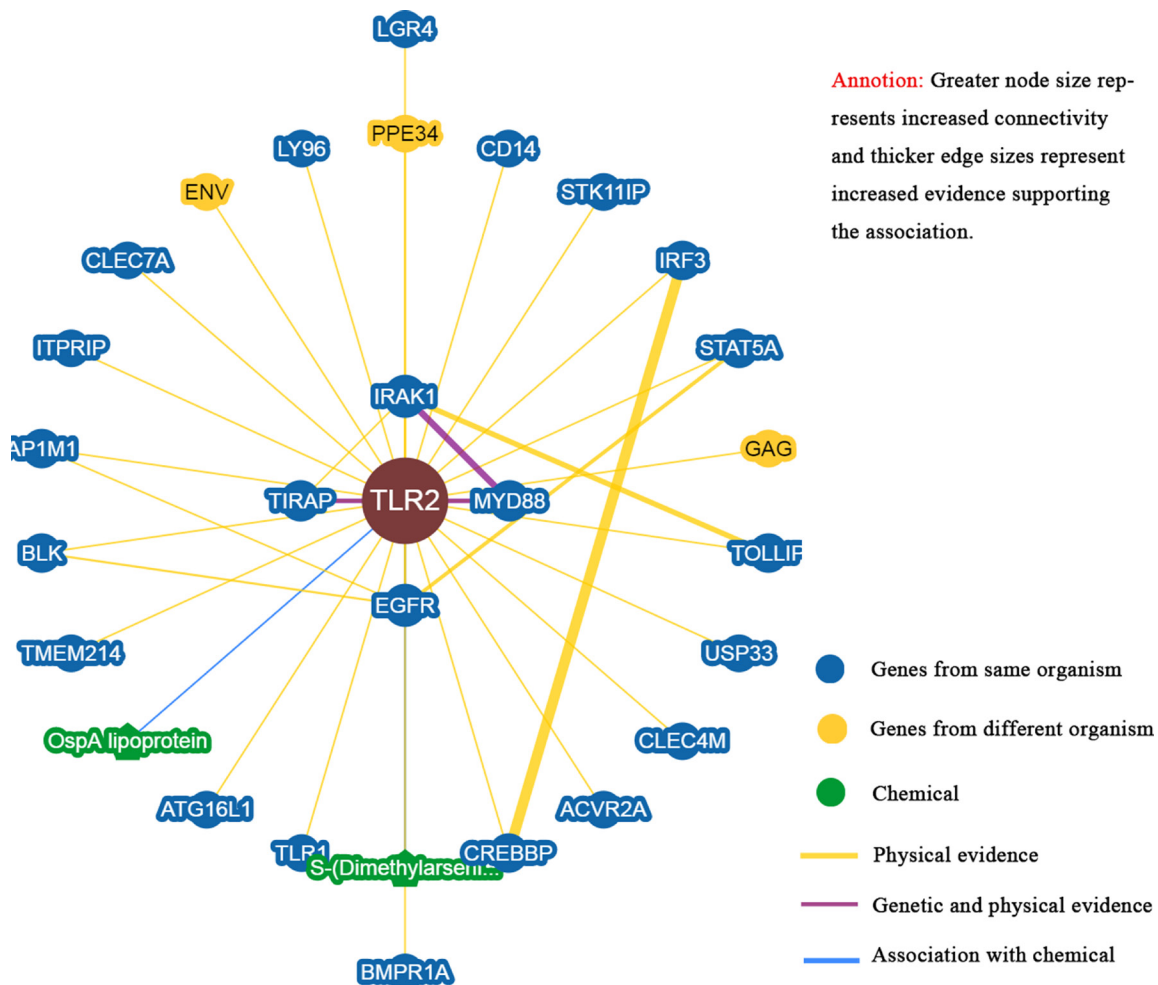


Figure 5: TLR2 interaction networks.

Identification of potential SL genes for immune genes

We identified the set of immune genes whose expression has significant difference between gene-mutated cancers and gene-wildtype cancers (Student's *t* test, FDR < 0.05, fold change > 1.5), and has significant difference between gene-mutated cancers and normal tissues (Student's *t* test, FDR < 0.05, fold change > 1.5). We identified potential SL genes for immune genes from the intersection of these two gene sets. To identify genes whose elevated expression is specifically related to other gene-mutated cancers, we believe that it is necessary to exclude as many genes as possible whose expression is significantly different between in gene-wildtype cancers and in normal tissues.

Comparison of drug sensitivity in cancer cell lines

We compared IC₅₀ values belonging to intestine tissue cancer cell lines between immune genes higher-expression-level and lower-expression-level cancer cell lines for compounds using Student's *t* test. We identified the compounds for which immune genes higher-expression-level and lower-expression-level cancer cell lines have significantly different IC₅₀ values using a threshold of *P*-value < 0.05.

Author contributions

JL conceived of and performed the research. QS performed data analyses, and wrote the manuscript. YC and SY performed data analyses. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This research was supported by High Performance Computing Center of China Pharmaceutical University.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by the National Natural Science Foundation of China (grant numbers: 81373482, 81373378).

REFERENCES

1. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, Jemal A. Colorectal cancer statistics, 2017. *CA Cancer J Clin.* 2017; 67:177–93. <https://doi.org/10.3322/caac.21395>.
2. Shah MA, Renfro LA, Allegra CJ, Andre T, de Gramont A, Schmoll HJ, Haller DG, Alberts SR, Yothers G, Sargent DJ. Impact of Patient Factors on Recurrence Risk and Time Dependency of Oxaliplatin Benefit in Patients With Colon Cancer: Analysis From Modern-Era Adjuvant Studies in the Adjuvant Colon Cancer End Points (ACCENT) Database. *J Clin Oncol.* 2016; 34:843–53. <https://doi.org/10.1200/JCO.2015.63.0558>.
3. Boland CR, Goel A. Prognostic Subgroups among Patients with Stage II Colon Cancer. *N Engl J Med.* 2016; 374:277–8. <https://doi.org/10.1056/Nejme1514353>.
4. Terzic J, Grivennikov S, Karin E, Karin M. Inflammation and colon cancer. *Gastroenterology.* 2010; 138:2101–14 e5. <https://doi.org/10.1053/j.gastro.2010.01.058>.
5. Arteaga CL, Baselga J. Impact of genomics on personalized cancer medicine. *Clin Cancer Res.* 2012; 18:612–8. <https://doi.org/10.1158/1078-0432.CCR-11-2019>.
6. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol.* 2011; 8:686–700. <https://doi.org/10.1038/nrgastro.2011.173>.
7. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012; 12:298–306. <https://doi.org/10.1038/nrc3245>.
8. Kirchnerberger S, Royston DJ, Boulard O, Thornton E, Franchini F, Szabady RL, Harrison O, Powrie F. Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. *J Exp Med.* 2013; 210:917–31.
9. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddock KF, Kim S, Davies SR, Wang S, Wang P, et al, and NCI CPTAC. Proteogenomic characterization of human colon and rectal cancer. *Nature.* 2014; 513:382–7. <https://doi.org/10.1038/nature13438>.
10. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* 2013; 41:D1228–33.
11. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, Thomson E, Monteiro R, Guimaraes P, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res.* 2014; 58:234–9. <https://doi.org/10.1007/s12026-014-8516-1>.
12. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, et al, and Inflamm and Host Response to Injury Large Scale Collab. Res. Program. A network-based analysis of systemic inflammation in humans. *Nature.* 2005; 437:1032–7. <https://doi.org/10.1038/nature03985>.
13. Kondelin J, Gylfe AE, Lundgren S, Tanskanen T, Hamberg J, Aavikko M, Palin K, Ristolainen H, Katainen R,

- Kaasinen E, Taipale M, Taipale J, Renkonen-Sinisalo L, et al. Comprehensive evaluation of protein coding mononucleotide microsatellites in microsatellite-unstable colorectal cancer. *Cancer Res.* 2017; 77:4078–88. <https://doi.org/10.1158/0008-5472.CAN-17-0682>.
14. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Taberero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer.* 2017; 17:268. <https://doi.org/10.1038/nrc.2017.24>.
 15. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Sonesson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* 2015; 21:1350–6. <https://doi.org/10.1038/nm.3967>.
 16. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013; 499:214–8. <https://doi.org/10.1038/nature12213>.
 17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
 18. Fehlker M, Huska MR, Jons T, Andrade-Navarro MA, Kemmner W. Concerted down-regulation of immune-system related genes predicts metastasis in colorectal carcinoma. *BMC Cancer.* 2014; 14:64. <https://doi.org/10.1186/1471-2407-14-64>.
 19. Catalan V, Gomez-Ambrosi J, Rodriguez A, Ramirez B, Izaguirre M, Hernandez-Lizoaian JL, Baixauli J, Marti P, Valenti V, Moncada R, Silva C, Salvador J, Fruhbeck G. Increased Obesity-Associated Circulating Levels of the Extracellular Matrix Proteins Osteopontin, Chitinase-3 Like-1 and Tenascin C Are Associated with Colon Cancer. *PLoS One.* 2016; 11:e0162189. <https://doi.org/10.1371/journal.pone.0162189>.
 20. Oh CD, Lu Y, Liang SD, Mori-Akiyama Y, Chen D, de Crombrughe B, Yasuda H. SOX9 Regulates Multiple Genes in Chondrocytes, Including Genes Encoding ECM Proteins, ECM Modification Enzymes, Receptors, and Transporters. *PLoS One.* 2014; 9:e107577. <https://doi.org/10.1371/journal.pone.0107577>.
 21. Ho GYF, Wang T, Zheng SQL, Tinker L, Xu JF, Rohan TE, Wassertheil-Smoller S, Xue XN, Augenlicht LH, Peters U, Phipps AI, Strickler HD, Gunter MJ, et al. Circulating Soluble Cytokine Receptors and Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers Prev.* 2014; 23:179–88. <https://doi.org/10.1158/1055-9965.EPI-13-0545>.
 22. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical And Powerful Approach To Multiple Testing. *J R Stat Soc B.* 1995; 57:289–300.
 23. Dannappel M, Vlantis K, Kumari S, Polykratis A, Kim C, Wachsmuth L, Eftychi C, Lin J, Corona T, Hermance N, Zelic M, Kirsch P, Basic M, et al. RIPK1 maintains epithelial homeostasis by inhibiting apoptosis and necroptosis. *Nature.* 2014; 513:90–4. <https://doi.org/10.1038/nature13608>.
 24. Jin Z, Liang F, Yang J, Mei W. hnRNP I regulates neonatal immune adaptation and prevents colitis and colorectal cancer. *PLoS Genet.* 2017; 13:e1006672. <https://doi.org/10.1371/journal.pgen.1006672>.
 25. Cabal-Hierro L, O'Dwyer PJ. TNF Signaling through RIP1 Kinase Enhances SN38-Induced Death in Colon Adenocarcinoma. *Mol Cancer Res.* 2017; 15:395–404.
 26. Long MJ, Wu FX, Li P, Liu M, Li X, Tang H. MicroRNA-10a targets CHL1 and promotes cell growth, migration and invasion in human cervical cancer cells. *Cancer Lett.* 2012; 324:186–96.
 27. Zhang XF, Li KK, Gao L, Li SZ, Chen K, Zhang JB, Wang D, Tu RF, Zhang JX, Tao KX, Wang G, Zhang XD. miR-191 promotes tumorigenesis of human colorectal cancer through targeting C/EBP β . *Oncotarget.* 2015; 6:4144–58. <https://doi.org/10.18632/oncotarget.2864>.
 28. Colbert LE, Petrova AV, Fisher SB, Pantazides BG, Madden MZ, Hardy CW, Warren MD, Pan Y, Nagaraju GP, Liu EA, Saka B, Hall WA, Shelton JW, et al. CHD7 expression predicts survival outcomes in patients with resected pancreatic cancer. *Cancer Res.* 2014; 74:2677–87.
 29. Savari S, Chandrashekar NK, Osman J, Douglas D, Bellamkonda K, Jönsson G, Juhas M, Greicius G, Pettersson S, Sjölander A. Cysteinyl leukotriene 1 receptor influences intestinal polyp incidence in a gender-specific manner in the ApcMin/+ mouse model. *Carcinogenesis.* 2016; 37:491–99.
 30. Peng YP, Xi CH, Zhu Y, Yin LD, Wei JS, Zhang JJ, Liu XC, Guo S, Fu Y, Miao Y. Altered expression of CD226 and CD96 on natural killer cells in patients with pancreatic cancer. *Oncotarget.* 2016; 7:66586–94. <https://doi.org/10.18632/oncotarget.11953>.
 31. Turcotte S, Gros A, Tran E, Lee CC, Wunderlich JR, Robbins PF, Rosenberg SA. Tumor-reactive CD8+ T cells in metastatic gastrointestinal cancer refractory to chemotherapy. *Clin Cancer Res.* 2014; 20:331–43. <https://doi.org/10.1158/1078-0432.CCR-13-1736>.
 32. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, Huso DL, Brancati FL, Wick E, McAllister F, Housseau F, Pardoll DM, Sears CL. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med.* 2009; 15:1016–22. <https://doi.org/10.1038/nm.2015>.
 33. The Cytoplasmic Sensor NLRC3 Inhibits mTOR Signaling in Tumors. *Cancer Discov.* 2017; 7:123. <https://doi.org/10.1158/2159-8290.CD-RW2017-001>.
 34. Karki R, Malireddi RK, Zhu Q, Kanneganti TD. NLRC3 regulates cellular proliferation and apoptosis to attenuate the development of colorectal cancer. *Cell Cycle.* 2017; 16:1243–51.

35. Salcedo R, Worschech A, Cardone M, Jones Y, Gyulai Z, Dai RM, Wang E, Ma W, Haines D, O'hUigin C, Marincola FM, Trinchieri G. MyD88-mediated signaling prevents development of adenocarcinomas of the colon: role of interleukin 18. *J Exp Med.* 2010; 207:1625–36.
36. Ma CI, Martin C, Ma Z, Hafiane A, Dai M, Lebrun JJ, Kiss RS. Engulfment protein GULP is regulator of transforming growth factor- β response in ovarian cells. *J Biol Chem.* 2012; 287:20636–51.
37. Wang R, Kang Y, Lohr CV, Fischer KA, Bradford CS, Johnson G, Dashwood WM, Williams DE, Ho E, Dashwood RH. Reciprocal regulation of BMF and BIRC5 (Survivin) linked to Eomes overexpression in colorectal cancer. *Cancer Lett.* 2016; 381:341–8. <https://doi.org/10.1016/j.canlet.2016.08.008>.
38. Wang X, Sun Q. TP53 mutations, expression and interaction networks in human cancers. *Oncotarget.* 2017; 8:624–43. <https://doi.org/10.18632/oncotarget.13483>.
39. Barlaam B, Cosulich S, Degorce S, Fitzek M, Giordanetto F, Green S, Inghardt T, Hennequin L, Hancox U, Lambert-van der Brempt C, Morgentin R, Pass S, Ple P, et al. Discovery of 9-(1-anilinoethyl)-2-morpholino-4-oxo-pyrido[1,2-a]pyrimidine-7-carboxamides as PI3K beta/delta inhibitors for the treatment of PTEN-deficient tumours. *Bioorg Med Chem Lett.* 2014; 24:3928–35. <https://doi.org/10.1016/j.bmcl.2014.06.040>.
40. Weigelt B, Warne PH, Lambros MB, Reis-Filho JS, Downward J. PI3K pathway dependencies in endometrioid endometrial cancer cell lines. *Clin Cancer Res.* 2013; 19:3533–44.
41. Sternberg CN, Eisen T, Tomczak P, Strahs AL, Esteves B, Berkenblit A, Motzer RJ. Tivozanib in patients treatment-naive for metastatic renal cell carcinoma: A subset analysis of the phase III TIVO-1 study. *Journal Of Clinical Oncology.* 2013; 31:4513–4513.
42. Hepgur M, Sadeghi S, Dorff TB, Quinn DI. Tivozanib in the treatment of renal cell carcinoma. *Biologics.* 2013; 7:139–48. <https://doi.org/10.2147/BTT.S32958>.
43. Kethiwale S, Borza CM, Lowe EW, Pozzi A, Meiler J. Discoidin domain receptor 1 (DDR1) kinase as target for structure-based drug discovery. *Drug Discov Today.* 2015; 20:255–61. <https://doi.org/10.1016/j.drudis.2014.09.025>.
44. Tsai J, Lee JT, Wang W, Zhang J, Cho H, Mamo S, Bremer R, Gillette S, Kong J, Haass NK, Sproesser K, Li L, Smalley KS, et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci USA.* 2008; 105:3041–46.
45. Tan J, Li Z, Lee PL, Guan P, Aau MY, Lee ST, Feng M, Lim CZ, Lee EY, Wee ZN, Lim YC, Karuturi RK, Yu Q. PDK1 signaling toward PLK1-MYC activation confers oncogenic transformation, tumor-initiating cell activation, and resistance to mTOR-targeted therapy. *Cancer Discov.* 2013; 3:1156–71.
46. Myers SH, Brunton VG, Unciti-Broceta A. AXL Inhibitors in Cancer: A Medicinal Chemistry Perspective. *J Med Chem.* 2016; 59:3593–608. <https://doi.org/10.1021/acs.jmedchem.5b01273>.
47. Godbersen JC, Humphries LA, Danilova OV, Kebbekus PE, Brown JR, Eastman A, Danilov AV. The Nedd8-Activating Enzyme Inhibitor MLN4924 Thwarts Microenvironment-Driven NF-kappa B Activation and Induces Apoptosis in Chronic Lymphocytic Leukemia B Cells. *Clin Cancer Res.* 2014; 20:1576–89. <https://doi.org/10.1158/1078-0432.CCR-13-0987>.
48. Liu KS, Ding WC, Wang SX, Liu Z, Xing GW, Wang Y, Wang YF. The heat shock protein 90 inhibitor SNX-2112 inhibits B16 melanoma cell growth *in vitro* and *in vivo*. *Oncol Rep.* 2012; 27:1904–10. <https://doi.org/10.3892/or.2012.1738>.
49. Hardwicke MA, Oleykowski CA, Plant R, Wang J, Liao Q, Moss K, Newlander K, Adams JL, Dhanak D, Yang J, Lai Z, Sutton D, Patrick D. GSK1070916, a potent Aurora B/C kinase inhibitor with broad antitumor activity in tissue culture cells and human tumor xenograft models. *Mol Cancer Ther.* 2009; 8:1808–17.
50. Hari DM, Leung AM, Lee JH, Sim MS, Vuong B, Chiu CG, Bilchik AJ. AJCC Cancer Staging Manual 7th Edition Criteria for Colon Cancer: Do the Complex Modifications Improve Prognostic Assessment? *Journal Of the American College Of Surgeons.* 2013; 217:181–90. <https://doi.org/10.1016/j.jamcollsurg.2013.04.018>.
51. McLornan DP, List A, Mufti GJ. Applying Synthetic Lethality for the Selective Targeting of Cancer. *N Engl J Med.* 2014; 371:1725–35. <https://doi.org/10.1056/Nejmra1407390>.
52. Hanaford AR, Archer TC, Price A, Kahlert UD, Maciaczyk J, Nikkha G, Kim JW, Ehrenberger T, Clemons PA, Dancik V, Seashore-Ludlow B, Viswanathan V, Stewart ML, et al. DiSCoVERing Innovative Therapies for Rare Tumors: Combining Genetically Accurate Disease Models with In Silico Analysis to Identify Novel Therapeutic Targets. *Clin Cancer Res.* 2016; 22:3903–14. <https://doi.org/10.1158/1078-0432.CCR-15-3011>.
53. Kim S, Takahashi H, Lin WW, Descargues P, Grivennikov S, Kim Y, Luo JL, Karin M. Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature.* 2009; 457:102–6. <https://doi.org/10.1038/nature07623>.
54. Lee H, Choi DW, Cho YB, Yun SH, Kim HC, Lee WY, Heo JS, Choi SH, Jung KU, Chun HK. Recurrence pattern depends on the location of colon cancer in the patients with synchronous colorectal liver metastasis. *Ann Surg Oncol.* 2014; 21:1641–6. <https://doi.org/10.1245/s10434-013-3477-5>.
55. Nakayama Y, Miura T, Hakamada K. Successful Resection of the Peritoneal Dissemination Recurrence of Colon Cancer, Including Metastasis to the Inguinal Hernia Sac--A Case Report. [Article in Chinese] *Gan To Kagaku Ryoho.* 2015; 42:1600–2.
56. Bansal K, Trinath J, Chakravorty D, Patil SA, Balaji KN. Pathogen-specific TLR2 Protein Activation Programs Macrophages to Induce Wnt-beta-Catenin Signaling. *J Biol Chem.* 2011; 286:37032–44. <https://doi.org/10.1074/jbc.M111.260414>.

57. Perez-Figueroa E, Torres J, Sanchez-Zauco N, Contreras-Ramos A, Alvarez-Arellano L, Maldonado-Bernal C. Activation of NLRP3 inflammasome in human neutrophils by *Helicobacter pylori* infection. *Innate Immun*. 2016; 22:103–12. <https://doi.org/10.1177/1753425915619475>.
58. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 2006; 34:D535-D9. <https://doi.org/10.1093/nar/gkj109>.
59. Picard C, von Bernuth H, Ghandil P, Chrabieh M, Levy O, Arkwright PD, McDonald D, Geha RS, Takada H, Krause JC, Creech CB, Ku CL, Ehl S, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine (Baltimore)*. 2010; 89:403–25. <https://doi.org/10.1097/MD.0b013e3181fd8ec3>.
60. Gray P, Dunne A, Brikos C, Jefferies CA, Doyle SL, O'Neill LAJ. MyD88 adapter-like (Mal) is phosphorylated by Bruton's tyrosine kinase during TLR2 and TLR4 signal transduction (vol 281, pg 10489, 2006). *Journal Of Biological Chemistry*. 2016; 291:26240-. <https://doi.org/10.1074/jbc.A116.508892>.
61. Kennedy CL, Najdovska M, Tye H, McLeod L, Yu L, Jarnicki A, Bhathal PS, Putoczki T, Ernst M, Jenkins BJ. Differential role of MyD88 and Mal/TIRAP in TLR2-mediated gastric tumorigenesis. *Oncogene*. 2014; 33:2540–6. <https://doi.org/10.1038/onc.2013.205>.
62. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30:207–10.
63. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG. Ensembl 2018. *Nucleic Acids Res*. 2018; 46:D754–D761.
64. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res*. 2015; 43:D1079.
65. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi M, Burdett T, Megy K, Pilicheva E, Rustici G, et al. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res*. 2015; 43:D1113–16.
66. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017; 45:D777–83.