

# Machine learning identifies two autophagy-related genes as markers of recurrence in colorectal cancer

Journal of International Medical Research  
48(10) 1–13

© The Author(s) 2020

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0300060520958808

[journals.sagepub.com/home/imr](https://journals.sagepub.com/home/imr)



Jianping Wu<sup>1</sup>, Sulai Liu<sup>2</sup>, Xiaoming Chen<sup>1</sup>,  
Hongfei Xu<sup>1</sup> and Yaoping Tang<sup>1</sup> 

## Abstract

**Objective:** Colorectal cancer (CRC) is the most common cancer worldwide. Patient outcomes following recurrence of CRC are very poor. Therefore, identifying the risk of CRC recurrence at an early stage would improve patient care. Accumulating evidence shows that autophagy plays an active role in tumorigenesis, recurrence, and metastasis.

**Methods:** We used machine learning algorithms and two regression models, univariable Cox proportion and least absolute shrinkage and selection operator (LASSO), to identify 26 autophagy-related genes (ARGs) related to CRC recurrence.

**Results:** By functional annotation, these ARGs were shown to be enriched in necroptosis and apoptosis pathways. Protein–protein interactions identified *SQSTM1*, *CASP8*, *HSP80AB1*, *FADD*, and *MAPK9* as core genes in CRC autophagy. Of 26 ARGs, *BAX* and *PARP1* were regarded as having the most significant predictive ability of CRC recurrence, with prediction accuracy of 71.1%.

**Conclusion:** These results shed light on prediction of CRC recurrence by ARGs. Stratification of patients into recurrence risk groups by testing ARGs would be a valuable tool for early detection of CRC recurrence.

## Keywords

Colorectal cancer, machine learning, regression, autophagy, autophagy-related gene, recurrence

Date received: 6 April 2020; accepted: 17 August 2020

<sup>1</sup>Hunan University of Science and Engineering, Yongzhou, China

<sup>2</sup>Department of Hepatobiliary Surgery, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha, China

## Corresponding author:

Yaoping Tang, Hunan University of Science and Engineering, 130 Yangzitang Road, Yongzhou 425100, China.

Email: [daiqilai1988@163.com](mailto:daiqilai1988@163.com)



## Introduction

Colorectal cancer (CRC) is one of the most common cancers and the third leading cause of cancer-related deaths worldwide.<sup>1</sup> The 5-year survival rate of CRC varies from 14% to 90%, based on the stage of CRC; patients diagnosed with localized cancers (stage I or II) usually have more favorable outcomes than patients with late-stage CRC.<sup>2</sup> Surgical resection of the tumor is the standard treatment for localized CRC but it cannot prevent recurrence after several years.<sup>3</sup> Because recurrence is preventable, research is needed to better predict the risk of CRC recurrence and stratify CRC patients into subgroups according to recurrence risk. A few studies in which CRC patients benefited from stratification into different adjuvant treatment groups according to molecular biomarkers have demonstrated the clinical utility of, and urgency to identify, molecular biomarkers in CRC.<sup>4-6</sup>

High-throughput screening and advanced bioinformatics analysis along with machine learning characterize the clinical utility of biomarkers in CRC. Two commercial assays (Epi proColon, Epigenomics AG, Berlin, Germany; and Cologuard, Exact Sciences, Madison, WI, USA) that are based on DNA methylation levels have been approved by the US Food and Drug Administration for CRC diagnosis and screening. However, most reported biomarkers in CRC are used for early cancer screening and therapeutic stratification.<sup>7-9</sup> There is a need to develop a predictive biomarker for CRC recurrence at all stages.

Autophagy is the process of transporting damaged, denatured, or aged proteins and organelles into lysosomes for digestion and degradation. On the one hand, autophagy prevents accumulation of toxic or carcinogenic proteins and organelles and inhibits cell carcinogenesis.<sup>10-13</sup> On the other hand, autophagy can be harmful because

the autophagy cells provide more nutrients and promote tumor growth when tumor cells form.<sup>14-17</sup> Because autophagy can regulate tumor formation, spread, metastasis, and energy metabolism, antitumor drugs based on the regulation of autophagy activity have the potential for clinical treatment.<sup>18-20</sup>

Depending on the way a lysosome accepts a substance to be degraded, autophagy can be classified into macroautophagy, microautophagy, and chaperone-mediated autophagy. Macroautophagy is the most common form of autophagy and is characterized by the formation of a cup-shaped bilayer membrane structure around the cytoplasmic component and the formation of an autophagosome.<sup>21-24</sup> The outer membrane and enzymatic fusion of autophagosomes form a monolayer membrane structure of autophagosomes, while the contents of the inner membrane and autophagosomes are digested.<sup>21-24</sup> This process is mediated by autophagy-related genes (ARGs).<sup>25-31</sup> These ARGs have been identified as direct or indirect participants in the autophagy process. Therefore, analysis of ARGs provides a comprehensive overview of the changes in autophagy in CRC. A number of studies have shown that these ARGs have important clinical implications for various types of cancer, including glioma, liver cancer, melanoma, and thyroid cancer.<sup>29,30,32,33</sup> In this study, we used two public datasets of CRC that included recurrent cases to explore the clinical utility of ARGs to predict CRC recurrence.

Because disease biomarkers demonstrate utility in disease diagnosis, prediction of therapeutic response, and monitoring of residual disease, machine learning has become a useful tool for biomarker discovery.<sup>34-36</sup> More importantly, machine learning is a powerful way to identify phenotypic features compared with conventional analysis of differentially expressed genes in cancers.

By using machine learning, we identified two ARGs with potential predictive ability for CRC recurrence; these genes were significantly associated with progression-free survival. Our results highlight the biological function of ARGs in CRC.

## Methods

### *Ethical approval*

This was a retrospective study using publicly available datasets; ethical approval was deemed unnecessary.

### *Data source and processing*

Two microarray datasets (GSE64857 and GSE28722) were downloaded from the Gene Expression Omnibus (GEO) database (<http://ncbi.nlm.nih.gov/geo/>). GSE64857 contained 81 samples of CRC patients with follow-up or recurrence information, which we used to identify ARG signatures. GSE28722 contained 129 samples of CRC patients with survival information, which we used to validate the prognostic capacity of ARG signatures in this cohort.<sup>37,38</sup> The annotation files of these microarrays were downloaded from the GEO website.

We acquired a list of ARGs from the human autophagy database (HADb; <http://autophagy.lu>), the first human autophagy database. As a public repository, it provides updated annotations of human genes that are reported to be related to autophagy. Information on 232 ARGs was obtained from this website.

After detecting the variance between samples by box-and-whisker plots, we used quantile normalization to remove unexpected variance within samples. After normalization, we conducted a principal component analysis (PCA) to explore genomic differences between samples with the top 1000 variant probes using the package “limma”<sup>39</sup> in R (<https://www.r-project.org/>).

### *Cox proportion and LASSO regressions*

To generate CRC recurrence-related ARG signatures, we performed two regressions on our datasets: univariable Cox proportion and least absolute shrinkage and selection operator (LASSO). Univariable Cox proportion regression was first conducted to filter out CRC recurrence survival-related ARGs. Once we obtained the shortlist of ARGs from the univariable proportion regression, we then performed the LASSO regression on these genes to test their association with CRC recurrence using the “glmnet” package in R.<sup>40</sup> The LASSO regression analysis identified eligible ARG signatures for the CRC recurrence risk score. We generated the CRC recurrence risk score using the ARG signatures and their corresponding coefficients discovered by the LASSO model. The risk score was calculated as follows: Risk score =  $\sum n_i \text{ARG}_i \times \beta_i$ , where  $n_i$  was the numbers of ARG signatures,  $i$  was each ARG signature, and  $\beta$  was the corresponding coefficient. After the risk score was determined, we categorized CRC samples from GSE64857 into two groups (predicted recurrence or predicted non-recurrence) according to their risk score. The performance of this risk score was evaluated by receiver-operator characteristic (ROC) curve using “survival receiver-operator characteristic (ROCR)” package.<sup>41</sup> The area under the curve (AUC) was calculated to assess the predictive accuracy.

### *Gene pathway analysis*

To identify the biological pathways of ARG signatures, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used for functional enrichment analysis by the R package “clusterProfiler”.<sup>42</sup> The significance of a pathway was determined by a false discovery rate (FDR) < 0.05. The Search Tool for the

Retrieval of Interacting Genes/Proteins (STRING) database (<https://string-db.org/>; version 10.0) was used to explore the protein–protein interactions (PPI) of the recurrence-associated ARGs.

### Survival analysis

For overall survival (OS) and progression-free survival (PFS) analysis, we used Kaplan–Meier plots to analyze the prognostic capacity of ARG signatures in CRC. We stratified all samples of GSE28722 into three groups (low, intermediate, high) based on their relative expression of ARG signatures. The log-rank test was performed to determine statistical significance, and a  $p$ -value  $< 0.05$  was considered significant. All survival analyses were conducted using the R packages “survival” and “survminer”.

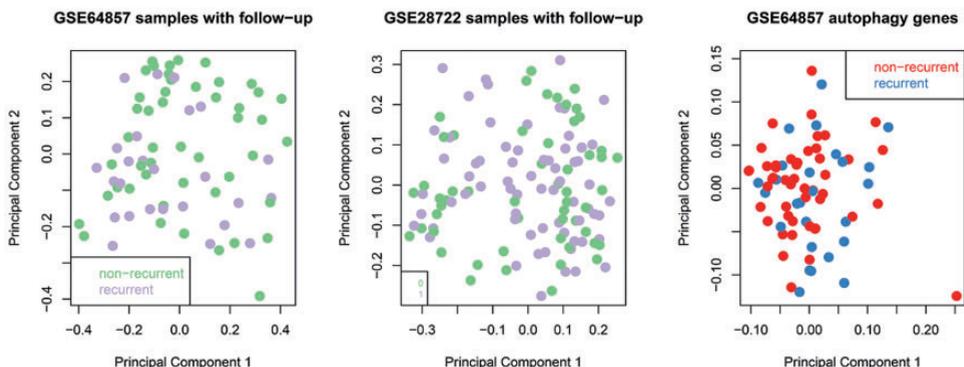
### Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed to investigate the pathways affected by the ARG signatures. We categorized samples of GSE64857 into two groups according to their risk score. GSEA was conducted using GSEA software provided by the Broad Institute (<http://software.broadinstitute.org/gsea/index.jsp>).<sup>43</sup> The hallmarks of genes were calculated if the signatures were enriched for CRC with a high or low risk score. Hallmark gene signatures with normalized enrichment score  $> 1$  and FDR  $< 0.05$  were considered significant.

## Results

### Download and cleaning of datasets for ARG signature identification

To generate the datasets for ARG signatures, we downloaded two datasets (GSE64857 and GSE28722) from the public domain (GEO). Dataset GSE64857 contained 81 samples with clinical information; 75 of these had clinical follow-up information, and 6 were excluded because they lacked recurrence information. Before conducting the bioinformatic analysis, we checked the variance within samples by performing quantile normalization for these two datasets. The boxplots of each sample from both datasets demonstrated that normalization removed the unwanted variance within samples (Figure 1). After normalization, we performed PCA to explore the variance of samples in terms of gene



**Figure 1.** Quantile normalization was used to remove the batch effect of data in two public datasets (GSE64857 and GSE28722) from the Gene Expression Omnibus.

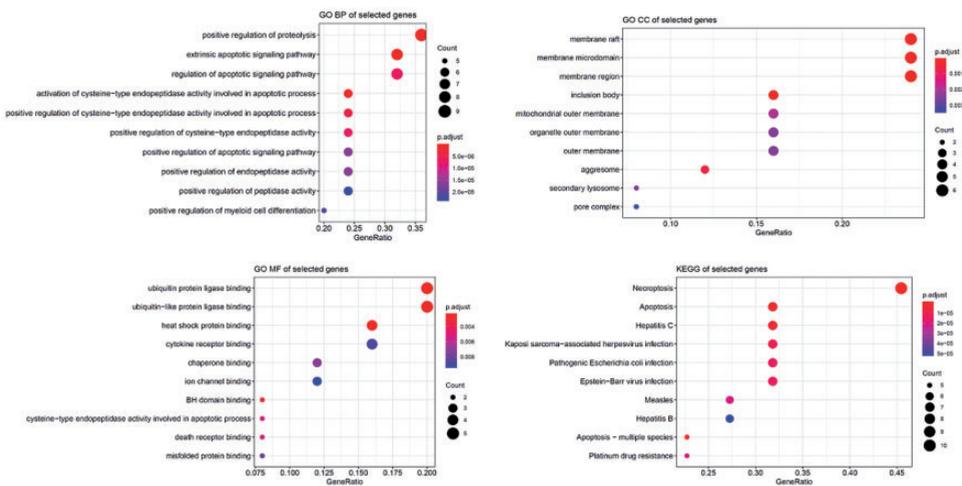
expression. Although the PCA plot of the top 1000 variable genes could not differentiate recurrent CRC from non-recurrent CRC, we observed separation within samples when we used the shortlist of ARGs (Figure 2). This result suggested that ARGs were potential biomarkers for CRC recurrence.

**Univariable Cox proportion regression identified ARGs for CRC recurrence**

To identify ARGs directly associated with CRC recurrence-free survival, we used univariable Cox proportion regression to filter all 232 ARGs from HADb (<http://autophagy.lu>). As the dataset GSE28722 used the Rosetta custom human 23K array, we included 283 probes targeting all 232 ARGs in the downstream analysis. Univariable Cox proportion regression showed that 26 ARGs were significantly associated with CRC recurrence-free survival (Table 1). Therefore, we used these 26 significant ARGs for molecular function annotation and LASSO analysis.

**Molecular function of ARG signatures**

To investigate the molecular function of these ARG signatures, we explored the associated enriched pathways in Gene Ontology (GO) terms (biological pathway, cellular component, and molecular function) and KEGG pathways. We used the ClusterProfiler package to identify the GO terms and KEGG pathways that ARG signatures enriched. For biological pathways of GO, the top three terms were “positive regulation of proteolysis,” “extrinsic apoptotic signaling pathway,” and “regulation of apoptotic signaling pathway.” The top three cellular component GO terms were “membrane raft,” “membrane microdomain,” and “membrane region.” The top three molecular function GO terms were “ubiquitin protein ligase binding,” “ubiquitin-like protein ligase binding,” and “heat shock protein binding.” The top three KEGG pathways enriched by the ARG signatures were “necroptosis,” “apoptosis,” and “hepatitis C” (Figure 3). The PPI network analysis demonstrated



**Figure 2.** Overview of biomarker discovery using two public datasets (GSE64857 and GSE28722) from the Gene Expression Omnibus. The two PCA plots on the left show the differences in samples for recurrent and non-recurrent colorectal cancer for the top 1000 variable genes. The right-hand panel shows the PCA plot indicating differences in samples by autophagy-related genes. PCA, principal component analysis.

**Table 1.** Univariable Cox proportion regression identified 26 ARGs associated with CRC recurrence.

Gene symbol	Beta	HR (95% CI for HR)	Wald test	p-value
<i>SQSTM1</i>	2.9	19 (3.3–110)	11	0.00097
<i>FOXO3</i>	1.7	5.7 (2–16)	11	0.0012
<i>FADD</i>	2.2	9.4 (2.2–40)	9	0.0027
<i>MAPK9</i>	–2.3	0.098 (0.02–0.47)	8.4	0.0038
<i>IFNG</i>	–0.76	0.47 (0.27–0.81)	7.5	0.0062
<i>EDEM1</i>	–1.9	0.16 (0.04–0.61)	7.1	0.0076
<i>PARP1</i>	–1.9	0.16 (0.039–0.64)	6.7	0.0097
<i>CASP8</i>	–1.4	0.25 (0.081–0.76)	5.9	0.015
<i>LAMP2</i>	1.5	4.5 (1.3–15)	6	0.015
<i>FKBP1A</i>	1.4	4.2 (1.3–14)	5.7	0.017
<i>CHMP2B</i>	–1.5	0.22 (0.061–0.79)	5.4	0.02
<i>ATG4D</i>	–2.6	0.075 (0.008–0.71)	5.1	0.024
<i>ITGA6</i>	–1.1	0.33 (0.13–0.88)	5	0.026
<i>DNAJB9</i>	–1.5	0.22 (0.057–0.85)	4.9	0.027
<i>BAX</i>	2.6	13 (1.3–140)	4.8	0.028
<i>KIAA0226</i>	–2.1	0.12 (0.018–0.8)	4.8	0.028
<i>RB1</i>	–1.2	0.31 (0.11–0.9)	4.6	0.031
<i>BAK1</i>	–1.2	0.31 (0.1–0.92)	4.4	0.035
<i>BNIP1</i>	–2.1	0.12 (0.017–0.86)	4.4	0.035
<i>C17orf88</i>	2.5	12 (1.2–120)	4.4	0.036
<i>EEF2</i>	1.7	5.4 (1.1–26)	4.4	0.036
<i>CASP4</i>	–1.3	0.28 (0.084–0.93)	4.3	0.038
<i>CASP4.1</i>	–1.3	0.28 (0.084–0.93)	4.3	0.038
<i>HSP90AB1</i>	1.8	6.2 (1–37)	4	0.046
<i>PRKARIA</i>	–1.4	0.25 (0.065–0.98)	4	0.046
<i>FAS</i>	–0.69	0.5 (0.26–0.99)	3.9	0.047

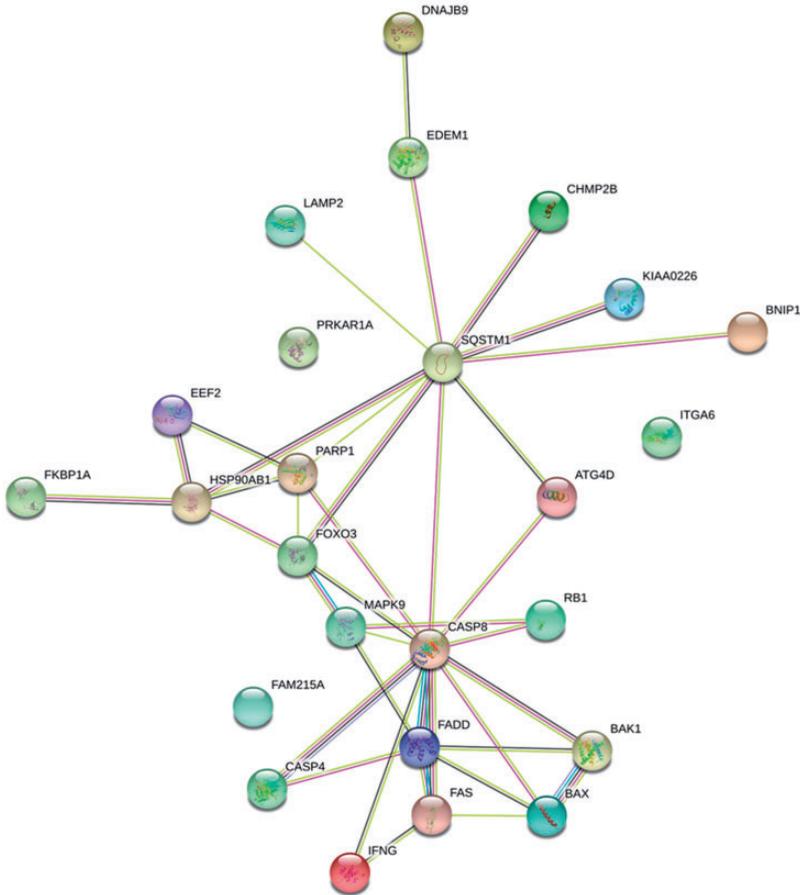
ARG, autophagy-associated gene; CRC, colorectal cancer; HR, hazard ratio; CI, confidence interval.

interactions in these ARG signatures. Of these, *SQSTM1*, *CASP8*, *HSP80AB1*, *FADD*, and *MAPK9* were the interaction center nodes, suggesting that these genes had broader interactions than other ARG signature genes. To further display ARG signature interactions, we performed a PPI analysis of the 26 ARGs using the STRING database. *SQSTM1*, *HSP90AB1*, and *CASP8* were identified as nodes in the PPI network of ARG signatures.

### Risk score stratification

To further test prediction of CRC recurrence by ARG signatures, we constructed a risk score prediction model by LASSO regression analysis. The 26 ARG signatures

shortlisted from the univariable Cox proportion analysis were subjected to LASSO regression analysis (Table 1). No genes retained a significant statistical association using the optimal lambda score in LASSO regression; therefore, we chose the minimum lambda score for LASSO regression. We identified two genes (*BAX* and *PARP1*) that were significantly associated with CRC recurrence (Figure 4) using the minimum lambda score in LASSO regression. Interestingly, the regression coefficients of these two genes were negative, suggesting that they had a protective effect on CRC recurrence. A CRC recurrence risk score was calculated for each patient by multiplying the regression coefficient by the

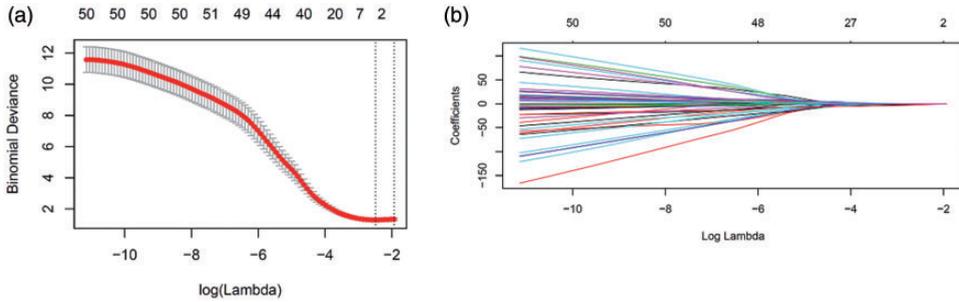


**Figure 3.** The enriched pathways of ARGs in CRC recurrence. The significance of annotated pathways is denoted by color and the number of genes enriched is indicated by the size of dots. (A) Top 10 biological processes enriched by ARGs; (B) top 10 cellular components of ARGs; (C) top 10 molecular functions of ARGs; and (D) top 10 KEGG pathways of ARGs. ARG, autophagy-related gene; CRC, colorectal cancer; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

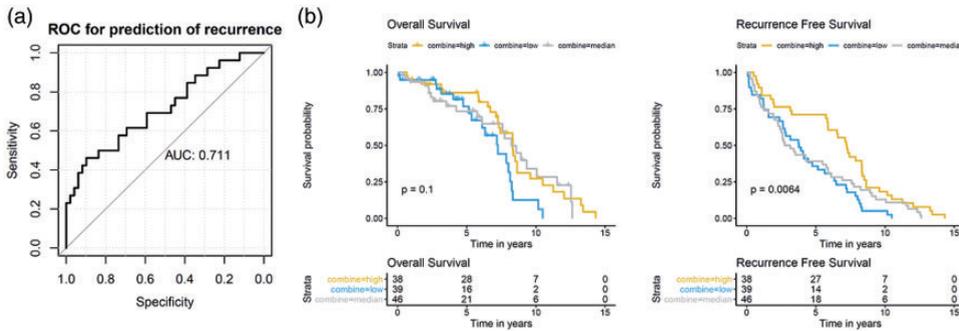
expression value of each gene. The CRC recurrence risk score =  $(208478\_s\_at, BAX) \times -1.3329 + (211833\_s\_at, BAX) \times -0.2936 + (208644\_at, PARP1) \times -2.2957$ , where 208478\_s\_at, 211833\_s\_at, and 208644\_at are the ARG gene probes for *BAX* and *PARP1*, respectively. To investigate the predictive accuracy of CRC recurrence risk score, we analyzed the ROC curve. The AUC of the CRC recurrence risk score ROC curve was 0.711 (Figure 5),

suggesting that a CRC recurrence risk score based on ARG signatures was a potential predictive biomarker.

To evaluate the prognostic ability of the CRC recurrence risk score, we performed a log-rank test on Kaplan–Meier curves of CRC survival. We categorized patients into three groups (high, low, and intermediate) based on *BAX* and *PARP1* gene expression. Patients with high or low expression of both *BAX* and *PARP1* were



**Figure 4.** The protein–protein interaction network showing the core genes in ARG signatures. Each node represents the gene signature, and the length of the lines indicates the degree of correlation between genes. ARG, autophagy-related gene.



**Figure 5.** Diagnostic and prognostic capacity of ARG signatures in CRC. (A) Coefficient of variance plot of LASSO regression. (B) Coefficient plot of LASSO regression results. (C) ROC curves showing the diagnostic capacity of ARG signatures for CRC recurrence; the area under the curve is 0.711. (D) Kaplan–Meier curves of CRC stratified by ARG signature using the log-rank test.

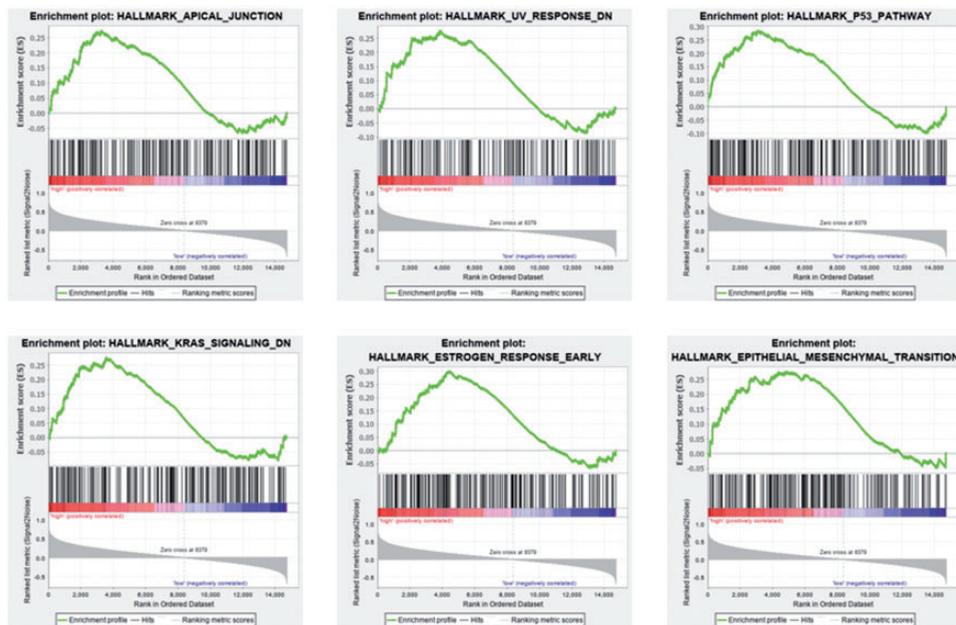
LASSO, least absolute shrinkage and selection operator; ARG, autophagy-related gene; CRC, colorectal cancer; ROC, receiver-operator characteristic; AUC, area under the curve.

assigned to the high or low group, respectively, whereas the rest of the patients were stratified into the intermediate group. Although CRC patients did not benefit from high expression of *BAX* and *PARP1* in overall survival, patients with high expression of *BAX* and *PARP1* had a significantly favorable outcome in recurrence-free survival (Figure 6). These results suggested that *BAX* and *PARP1* are potential prognostic biomarkers for CRC recurrence. Taken together, our results showed that *BAX* and *PARP1*

from ARG signatures could be utilized for CRC recurrence risk stratification.

### Hallmarks of ARG signatures

To validate the enriched pathways associated with ARG signatures, we performed GSEA for the prognostic markers *BAX* and *PARP1*. GSEA demonstrated six hallmark pathways: “estrogen response early,” “p53 pathway,” “Kras signaling DN,” “apical junction,” “epithelial–mesenchymal transition,” and “UV response DN,” that



**Figure 6.** Validation of ARG signatures in an independent cohort. (A) Boxplots showing the differential expression of *BAX* and *PARP1* between recurrent and non-recurrent CRC. (B) ROC curves showing the accuracy of the two ARG signatures.

ARG, autophagy-related gene; CRC, colorectal cancer; ROC, receiver-operator characteristic; AUC, area under the curve.

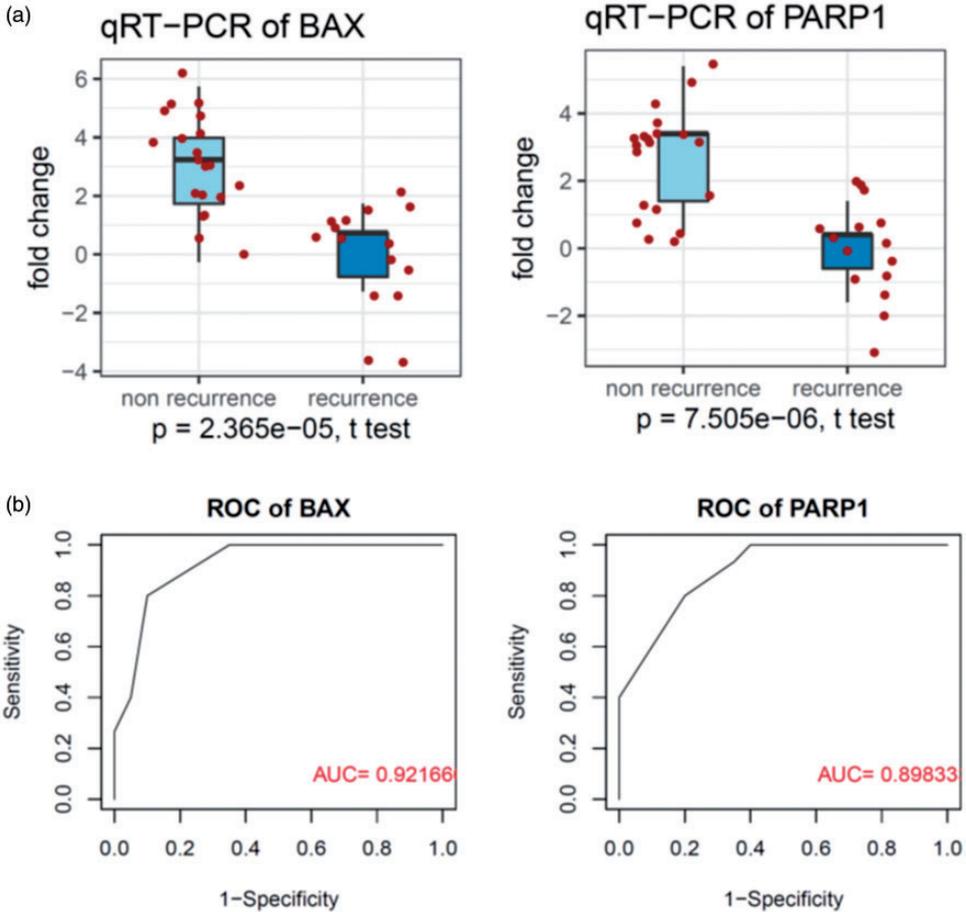
were significantly associated with CRC patients having high expression of *BAX* and *PARP1* (Figure 7).

## Discussion

To accurately predict the prognosis of CRC, many cohorts have been studied to establish gene expression characteristics.<sup>44,45</sup> A meta-analysis has been performed to assess the clinical utility of several published prognostic gene expression profiles in CRC.<sup>46</sup> Although most published signatures show a statistically significant association with prognosis, their accuracy in classifying independent tumor samples into high- and low-risk groups remains limited. Therefore, more robust and accurate gene expression profiles are needed to predict prognosis. Here, we established CRC recurrence gene signatures

by applying two supervised machine learning approaches. These recurrence signatures based on ARGs could predict CRC recurrence with moderate accuracy.

The role of autophagy in tumors is well known, and the role of autophagy in the development and treatment of CRC has been reported previously.<sup>47,48</sup> However, the clinical significance of ARGs, especially their prognostic effects in CRC, has not been extensively investigated. Moreover, there is no comprehensive analysis of the prognostic significance of ARGs. Therefore, in this study, we first analyzed the expression levels of ARGs and subsequently examined their prognostic value. Finally, we established a model using prognostic ARGs. Given the clinical significance of these prognostic ARGs in CRC, they may provide new directions for clinical



**Figure 7.** GSEA showing six hallmark pathways (apical junction, UV response DN, p53 pathway, Kras signaling DN, estrogen response early, and epithelial–mesenchymal transition) that were enriched in the high score ARG signature group.

GSEA, gene set enrichment analysis; ARG, autophagy-related gene.

CRC treatment if a CRC recurrence score can be calculated for each patient.

We evaluated the prognostic value of all 232 ARGs in CRC. By performing a univariate Cox analysis, we identified 26 recurrence-associated ARGs, suggesting that autophagy plays a crucial role in the development of CRC and may affect the prognosis of CRC patients. Because ARGs may have multiple functions in addition to autophagy, we also performed GO and KEGG pathway analysis and showed

that most enrichment pathways were autophagy-related pathways. These results revealed that autophagy is a key factor in CRC recurrence.

To more accurately assess prognosis of CRC patients, LASSO regression was used to analyze the prognostic significance of ARGs. As a result of LASSO regression, the number of ARGs was reduced to two genes, including three probes (208478\_s\_at, 211833\_s\_at, and 208644\_at). These ARGs were combined into one prognostic

assessment model. From the ROC curve, we found that the prognostic model had an AUC >0.7, indicating moderate efficacy in evaluating the prognosis and recurrence in patients with CRC. This moderate accuracy suggested there might be other confounding factors participating in CRC recurrence. The establishment of these prognostic biomarkers confirms the role of autophagy in the development of CRC and patient prognosis. Although BAX and PARP1 are widely recognized as key mediators in cell death and apoptosis, their roles in autophagy remain unclear for cancer development. Activation of PARP1 promotes expression of other autophagic pathway genes.<sup>49</sup> Increased ARG expression is caused by PARP1-mediated poly ADP-ribosylation that restrains FoxO3a in nuclei. In CRC, BAX is recognized as a negative regulator in autophagy; a few autophagy genes were inversely correlated with BAX.<sup>50</sup>

This proof-of-concept study has some limitations. The ARG signatures and recurrence-predicting genes identified in this bioinformatics study need to be validated in external cohorts. Additionally, we did not compare the accuracy of our ARG signatures with other existing biomarkers (genes, DNA methylation, or metabolites). Combined with other protein biomarkers, our ARG signatures might better predict recurrence. A duplex PCR-based assay targeting these two ARGs signatures is under development and will be tested for clinical utility soon.

In conclusion, we established an ARG signature that could predict CRC recurrence with moderate accuracy. The risk score based on ARG signature is a potential prognostic biomarker for CRC recurrence, but it requires clinical validation through further studies.

### Declaration of conflicting interest

The authors declare that there is no conflict of interest.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### ORCID iD

Yaoping Tang  <https://orcid.org/0000-0002-0763-2540>

### Supplemental Material

Supplemental material for this article is available online.

### References

1. McGuire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv Nutr* 2016; 7: 418–419.
2. Coleman MP, Quaresma M, Berrino F, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008; 9: 730–756.
3. Renehan AG, Malcomson L, Emsley R, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol* 2016; 17: 174–183.
4. Tie J, Cohen JD, Wang Y, et al. Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. *JAMA Oncol* 2019; 5: 1710–1717.
5. Tie J, Semira C and Gibbs P. Circulating tumor DNA as a biomarker to guide therapy in post-operative locally advanced rectal cancer: the best option? *Expert Rev Mol Diagn* 2018; 18: 1–3.
6. Tabernero J, Lenz HJ, Siena S, et al. Analysis of circulating DNA and protein biomarkers to predict the clinical activity of regorafenib and assess prognosis in patients with metastatic colorectal cancer: a retrospective, exploratory analysis of the CORRECT trial. *Lancet Oncol* 2015; 16: 937–948.

7. Sveen A, Kopetz S and Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat Rev Clin Oncol* 2020; 17: 11–32. doi: 10.1038/s41571-019-0241-1.
8. Marcuello M, Vymetalkova V, Neves RPL, et al. Circulating biomarkers for early detection and clinical management of colorectal cancer. *Mol Aspects Med* 2019; 69: 107–122.
9. Robertson DJ and Ladabaum U. Opportunities and challenges in moving from current guidelines to personalized colorectal cancer screening. *Gastroenterology* 2019; 156: 904–917.
10. Boga JA, Caballero B, Potes Y, et al. Therapeutic potential of melatonin related to its role as an autophagy regulator: A review. *J Pineal Res* 2019; 66: e12534.
11. Guillaume JD, Celano SL, Martin KR, et al. Determining the impact of metabolic nutrients on autophagy. *Methods Mol Biol* 2019; 1862: 151–162.
12. Liang ZG, Lin GX, Yu BB, et al. The role of autophagy in the radiosensitivity of the radioresistant human nasopharyngeal carcinoma cell line CNE-2R. *Cancer Manag Res* 2018; 10: 4125–4134.
13. Zhao Y, Onda K, Sugiyama K, et al. Antitumor effects of arsenic disulfide on the viability, migratory ability, apoptosis and autophagy of breast cancer cells. *Oncol Rep* 2019; 41: 27–42.
14. Jiao YN, Wu LN, Xue D, et al. *Marsdenia tenacissima* extract induces apoptosis and suppresses autophagy through ERK activation in lung cancer cells. *Cancer Cell Int* 2018; 18: 149.
15. Sun T, Liu H and Ming L. Multiple roles of autophagy in the sorafenib resistance of hepatocellular carcinoma. *Cell Physiol Biochem* 2017; 44: 716–727.
16. Wu Y, Liu X, Qin Z, et al. Low-frequency ultrasound enhances chemotherapy sensitivity and induces autophagy in PTX-resistant PC-3 cells via the endoplasmic reticulum stress-mediated PI3K/Akt/mTOR signaling pathway. *Oncotargets Ther* 2018; 11: 5621–5630.
17. Zhu J, Zhao B, Xiong P, et al. Curcumin induces autophagy via inhibition of Yes-associated protein (YAP) in human colon cancer cells. *Med Sci Monit* 2018; 24: 7035–7042.
18. Duan X, Chen B, Cui Y, et al. Ready player one? Autophagy shapes resistance to photodynamic therapy in cancers. *Apoptosis* 2018; 23: 587–606.
19. Han Y, Fan S, Qin T, et al. Role of autophagy in breast cancer and breast cancer stem cells (Review). *Int J Oncol* 2018; 52: 1057–1070.
20. Ianniciello A, Rattigan KM and Helgason GV. The ins and outs of autophagy and metabolism in hematopoietic and leukemic stem cells: food for thought. *Front Cell Dev Biol* 2018; 6: 120.
21. Feldmann A, Bekbulat F, Huesmann H, et al. The RAB GTPase RAB18 modulates macroautophagy and proteostasis. *Biochem Biophys Res Commun* 2017; 486: 738–743.
22. Han Q, Deng Y, Chen S, et al. Downregulation of ATG5-dependent macroautophagy by chaperone-mediated autophagy promotes breast cancer cell metastasis. *Sci Rep* 2017; 7: 4759.
23. Pajares M, Jimenez-Moreno N, Garcia-Yague AJ, et al. Transcription factor NFE2L2/NRF2 is a regulator of macroautophagy genes. *Autophagy* 2016; 12: 1902–1916.
24. Wang C, Wang H, Zhang D, et al. Phosphorylation of ULK1 affects autophagosome fusion and links chaperone-mediated autophagy to macroautophagy. *Nat Commun* 2018; 9: 3492.
25. Bednarczyk M, Muc-Wierzgon M, Waniczek D, et al. Autophagy-related gene expression in colorectal cancer patients. *J Biol Regul Homeost Agents* 2017; 31: 923–927.
26. Cao QH, Liu F, Yang ZL, et al. Prognostic value of autophagy related proteins ULK1, Beclin 1, ATG3, ATG5, ATG7, ATG9, ATG10, ATG12, LC3B and p62/SQSTM1 in gastric cancer. *Am J Transl Res* 2016; 8: 3831–3847.
27. Chen D, Chen J, Guo Y, et al. Cinobufacini promotes apoptosis of bladder cancer cells by influencing the expression of autophagy-related genes. *Oncol Lett* 2018; 15: 7104–7110.
28. Li WL, Xiong LX, Shi XY, et al. IKKbeta/NFkappaBp65 activated by interleukin-13

- targets the autophagy-related genes LC3B and beclin 1 in fibroblasts co-cultured with breast cancer cells. *Exp Ther Med* 2016; 11: 1259–1264.
29. Lin P, He Y, Wen DY, et al. Comprehensive analysis of the clinical significance and prospective molecular mechanisms of differentially expressed autophagy-related genes in thyroid cancer. *Int J Oncol* 2018; 53: 603–619.
  30. Ma Y, Zhang Y, Zhao Y, et al. Expression of autophagy-related genes in cerebrospinal fluid of patients with tuberculous meningitis. *Exp Ther Med* 2018; 15: 4671–4676.
  31. Zheng LQ, Li SY and Li CX. Expression profiling analysis of autophagy-related genes in perineural invasion of cutaneous squamous cell carcinoma. *Oncol Lett* 2018; 15: 4837–4848.
  32. Zhang H, Lu X, Wang N, et al. Autophagy-related gene expression is an independent prognostic indicator of glioma. *Oncotarget* 2017; 8: 60987–61000.
  33. Zhu FX, Wang XT, Zeng HQ, et al. A predicted risk score based on the expression of 16 autophagy-related genes for multiple myeloma survival. *Oncol Lett* 2019; 18: 5310–5324.
  34. Specogna AV and Sinicrope FA. Defining colon cancer biomarkers by using deep learning. *Lancet* 2020; 395: 314–316.
  35. Courtiol P, Maussion C, Moarii M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med* 2019; 25: 1519–1525.
  36. Ma C, Nguyen HPT, Luwor RB, et al. A comprehensive meta-analysis of circulation miRNAs in glioma as potential diagnostic biomarker. *PLoS One* 2018; 13: e0189452.
  37. Zhu H, Yu J, Zhu H, et al. Identification of key lncRNAs in colorectal cancer progression based on associated protein-protein interaction analysis. *World J Surg Oncol* 2017; 15: 153.
  38. Loboda A, Nebozhyn MV, Watters JW, et al. EMT is the dominant program in human colon cancer. *BMC Med Genomics* 2011; 4: 9.
  39. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
  40. Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; 33: 1–22.
  41. Sing T, Sander O, Beerenwinkel N, et al. ROCr: visualizing classifier performance in R. *Bioinformatics* 2005; 21: 3940–3941.
  42. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284–287.
  43. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102: 15545–15550.
  44. Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. *J Clin Oncol* 2011; 29: 4611–4619.
  45. Venook AP, Niedzwiecki D, Lopatin M, et al. Biologic determinants of tumor recurrence in stage II colon cancer: validation study of the 12-gene recurrence score in cancer and leukemia group B (CALGB) 9581. *J Clin Oncol* 2013; 31: 1775–1781.
  46. Sanz-Pamplona R, Berenguer A, Cordero D, et al. Clinical value of prognosis gene expression signatures in colorectal cancer: a systematic review. *PLoS One* 2012; 7: e48877.
  47. Kinsey CG, Camolotto SA, Boespflug AM, et al. Protective autophagy elicited by RAF->MEK->ERK inhibition suggests a treatment strategy for RAS-driven cancers. *Nat Med* 2019; 25: 620–627.
  48. Wei J, Long L, Yang K, et al. Autophagy enforces functional integrity of regulatory T cells by coupling environmental cues and metabolic homeostasis. *Nat Immunol* 2016; 17: 277–285.
  49. Wang C, Xu W, Zhang Y, et al. PARP1 promote autophagy in cardiomyocytes via modulating FoxO3a transcription. *Cell Death Dis* 2018; 9: 1047.
  50. Gil J, Ramsey D, Szmida E, et al. The BAX gene as a candidate for negative autophagy-related genes regulator on mRNA levels in colorectal cancer. *Med Oncol* 2017; 34: 16.