Development and validation of an immune-related gene pairs signature in colorectal cancer

Jianping Wu\textsuperscript{a}, Ying Zhao\textsuperscript{a}, Juanwen Zhang\textsuperscript{a}, Qianxia Wu\textsuperscript{a}, and Weilin Wang\textsuperscript{b,c}

\textsuperscript{a}Department of Clinical Laboratory, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; \textsuperscript{b}Key Laboratory of Precision Diagnosis and Treatment for Hepatobiliary and Pancreatic Tumor of Zhejiang Province, Division of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; \textsuperscript{c}State Key Laboratory & Collaborative Innovation Center for Diagnosis and Treatment of Infectious Disease, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China

**ABSTRACT**

Although the outcome of colorectal cancer (CRC) patients has improved significantly with the recent implementation of annual screening programs, reliable prognostic biomarkers are still needed due to the disease heterogeneity. Increasing pieces of evidence revealed an association between immune signature and CRC prognosis. Thus, we aim to build a robust immune-related gene pairs (IRGPs) signature that can estimate prognosis for CRC. Gene expression profiles and clinical information of CRC patients were collected from six public cohorts, divided into training cohort (n = 565) and five independent validation cohorts (n = 572, 290, 90 177 and 68, respectively). Within 1534 immune genes, a 19 IRGPs signature consisting of 36 unique genes was constructed which was significantly associated with the survival. In the validation cohorts, the IRGPs signature significantly stratified patients into high- vs low-risk groups in terms of prognosis across and within subpopulations with early stages disease and was prognostic in univariate and multivariate analyses. Several biological processes, including response to bacterium, were enriched among genes in the IRGPs signature. Macrophage M2 and mast cells were significantly higher in the high-risk risk group compared with the low-risk group. The IRGPs signature achieved a higher accuracy than commercialized multigene signatures for estimation of survival. When integrated with clinical factors such as sex and stage, the composite clinical and IRGPs signature showed improved prognostic accuracy relative to IRGPs signatures alone. In short, we developed a robust IRGPs signature for estimating prognosis in CRC, including early-stage disease, providing new insights into the identification of CRC patients with a high risk of mortality.

**Introduction**

Colorectal cancer (CRC) is a leading cause of cancer-related death worldwide.\textsuperscript{1,2} Nearly 1.4 million people are diagnosed as new cases of colorectal cancer every year.\textsuperscript{3} Although new tests and treatments have been achieved for the management of CRC, the 5-year survival rate is only approximately 55%.\textsuperscript{4} Targeted therapies including anti-EGFR antibodies, anti-angiogenic and immunotherapy agents may extend survival in selected patients.\textsuperscript{5} Surgical remains the first priority means of curative treatment. However, a proportion of patients will suffer local recurrences and remote metastases after surgery. Meanwhile, patients with equal clinical or pathologic conditions show unpredictable clinical outcomes, even treated similarity.\textsuperscript{6} The patients’ genetic heterogeneity contributes most for the inherent clinical diversity.\textsuperscript{7}

Biomarkers that can reliably estimate disease prognosis and patient survival would have tremendous value in guiding the management of CRCs.\textsuperscript{8,9} For prognosis biomarkers, researchers have investigated the possibility of stratifying patients with CRC based on gene expression signatures, and they have built multigene-expression signatures that can be used to stratify high-risk subgroup.\textsuperscript{10–12} Unfortunately, none has been incorporated into routine clinical practice owing to issues such as overfitting on small discovery data sets and lack of sufficient validation.\textsuperscript{14} The availability of public, large-scale gene expression data sets brings the opportunity to identify potentially more reliable CRC biomarkers.\textsuperscript{15–17} To use all this information effectively, however, the diversity of data also represents a daunting challenge. Gene expression levels sequenced by traditional approaches require suitable normalization and technical biases across sequencing platforms.\textsuperscript{18} Instead, researchers have proposed new methods to eliminate the limitations for data processing, such as normalization and scaling based on the relative ranking of gene expression levels, and have produced robust outcomes in various studies.\textsuperscript{19–21}

There is increasing evidence that the immune system plays a decisive role during cancer initiation and progression.\textsuperscript{22,23} Immune checkpoints tightly regulated immune function to maintain self-tolerance and minimize tissue destruction upon immune reaction in the peripheral tissues.\textsuperscript{24,25} Several
immune checkpoints have been discovered, and these immune checkpoints form therapeutic targets. The two immune checkpoint pathways that have received the most attention include programmed cell death 1 (PD-1) and cytotoxic T-lymphocyte associated antigen 4 (CTLA-4). Recent immunotherapies targeting specific immune checkpoints such as programmed death-1 (PD-1) or PD-1 ligand 1 (PD-L1) have demonstrated a remarkable, durable response in CRC.\(^{26,27}\) Regarding their prognostic potential in CRC, with poor prognosis, while plasma cell and Table S1). CIT (GSE39582) cohort was used as the training and independent validation data sets. Theprognostic index (ICPI) as in training cohort and similar prognostic value in 1 were independent predic-

mast cell and Figure S7; mean C-index, 0.89 vs 0.77 in CIT factors as the validation and evaluation of the IRGPI. Stage, the IRGPI remained an independent prognostic factor for the clinical and pathological factors such as age, sex and tumor stage (I and II) CRC only, the IRGPI remained highly prognostic in terms of RFS (Figure 1(c), HR, 2.43 [1.27–4.66]; \(P = 5.82 \times 10^{-3}\)). Considering early-stage, the IRGPI remained highly prognostic for the meta-validation cohorts (Figure 1(f) and Figure S4, HR, 2.51 [1.66–3.79]; \(P = 5.69 \times 10^{-5}\)). When considering patients in independent validation cohorts, the high-risk group had significantly poor RFS than patients in the low-risk group. (Figure S3). When compared with a clinically applicable and commercialized biomarker, the IRGPI achieved a higher C-index compared with Oncotype Dx Colon Cancer\(^{30}\) in training and independent validation data sets (Figure S5).

**Results**

**Construction of IRGPs and its prognostic value**

A total of 1,762 patients with CRC were included in the analysis (Table 1 and Table S1). CIT (GSE39582) cohort was used as the discovery and training data set. In total, 1,811 immune-related genes (IRGs) from the ImmPort database (accessed on 1/30/2018), including 17 categories. 474 IRGs, which were measured on all platforms and fulfilled the criteria on the training set (Median absolute deviation, MAD > 0.5). From these 474 IRGs, 112,101 immune-related gene pairs (IRGPs) were constructed. After removing IRGPs with relatively small variation (MAD = 0), 215 IRGPs were left and selected as initial candidate IRGPs. Then we defined IRGP index (IRGPI) using Lasso Cox proportional hazards regression on the training set and selected 19 IRGPs in the final model. The IRGPI consisted of 36 unique IRGs (Table S2). The optimal cut-off of the IRGPI for prediction of CRC prognosis was determined to be 0.704 using time-dependent ROC curve analysis (Figure S2). The IRGPI significantly divided patients into low- and high-risk groups in terms of relapse-free survival (RFS) (Figure 1(a-b) and Table S3, HR, 7.43 [4.23–13.07]; \(P = 3.33 \times 10^{-16}\) and overall survival (OS) (Figure 2(a-b), HR, 2.64 [1.76–3.94]; \(P = 9.34 \times 10^{-7}\)) for the training cohort. When considering patients with early-stage (I and II) CRC only, the IRGPI remained highly prognostic in terms of RFS (Figure 1(c), HR, 6.41 [3.26–12.63]; \(P = 7.13 \times 10^{-10}\)) for the training data set. To further investigate whether the IRGPI could serve as an independent predictor of prognosis, uni- and multivariate Cox regression analyses were applied to the training cohort. After adjusting for the clinical and pathological factors such as age, sex and tumor stage, the IRGPI remained an independent prognostic factor (Table S4 and Table S5).

**Validation and evaluation of the IRGPI**

To determine whether the IRGPI had similar prognostic value in different populations, we applied the same formula to five different cohorts from the TCGA, Jorissen, De Sousa, Smith and Kirzin databases as external validation sets. As expected, the IRGPI significantly stratified patients in terms of RFS (Table S3 and Figure 1(d-e), HR, 2.23 [1.71–2.90]; \(P = 9.37 \times 10^{-10}\)) in meta-validation data sets. On the TCGA cohort, high-risk group had significantly worse OS than the low-risk patients (Figure 2(c-d), HR, 2.43 [1.27–4.66]; \(P = 5.82 \times 10^{-3}\)). Considering early-stage, the IRGPI remained highly prognostic for the meta-validation cohorts (Figure 1(f) and Figure S4, HR, 2.51 [1.66–3.79]; \(P = 5.69 \times 10^{-5}\)). When considering patients in independent validation cohorts, the high-risk group had significantly poor RFS than patients in the low-risk group. (Figure S3). When compared with a clinically applicable and commercialized biomarker, the IRGPI achieved a higher C-index compared with Oncotype Dx Colon Cancer\(^{30}\) in training and independent validation data sets (Figure S5).

**In silico functional analysis of the immune-related risk model**

In order to gain new insights into the biological role of the obtained risk groups, we carried out immune infiltration and GO analysis. For immune infiltration, such as plasma cell, macrophage m2, T cells CD4 memory resting and T cells CD8, was enriched in training cohort (Figure 3(a)). We found that the percentages of macrophage m2, mast cell infiltration and plasma cell were significantly different between IRGPI risk groups. Previous studies have reported the association of macrophage M2,\(^{31}\) mast cell\(^{32}\) with poor prognosis, while plasma cell\(^{33}\) as the indicator of better prognosis. Among high-risk group, macrophage M2 and mast cell infiltration showed significantly higher than low-risk group, while plasma cell significantly enriched in low-risk group (Figure 3(b)). Furthermore, the risk groups specific immune cell subsets infiltration was also validated in meta-validation cohorts (Figure 3(c)). GO analysis revealed that the genes relevant to the IRGPI in training cohort were mostly involved in the immune response (Table S6 and Figure S6; GO terms, such as defense response, immune response, immune system process, inflammatory response and cytokine activity).

**Integrated prognostic index by combining the IRGPI with clinical factors**

In multivariate analysis (Table S4), sex, tumor stage and IRGPI were independent predictors of prognosis, suggesting their covariate value. To further improve accuracy, we fit a Cox proportional hazards regression model using sex, tumor stage and IRGPI score in training cohort and derived an immune-clinical prognostic index (ICPI) as (1.085 × sex) + (0.737 × stage) + (2.464 × IRGPI score). All three factors showed statistical significance in the Cox model (\(P = 0.000155, 0.000154, 2.49 \times 10^{-14}\), respectively) and positive correlation with the final risk, indicating the improvement of risk prediction. The optimal cut off of the IRGPI for stratifying patients was determined to be 0.871 using time-dependent ROC curve analysis in the training data set. Significantly improved estimation of survival was achieved by the continuous form of ICPI relative to IRGPI (Figure 4 and Figure S7; mean C-index, 0.89 vs 0.77 in CIT cohort).

**Discussion**

In this study, we developed an immune-related gene pairs signature reflecting distinct biological processes (immune infiltration)
to predict prognosis in colorectal cancer. The prognostic signature associated with tumor immune microenvironment may open up a new perspective for developing a novel predictive biomarker and improving CRC patient management in the era of immunotherapy. Most genes involved in our immune signature were antimicrobials, cytokines and cytokine receptors, which play key roles in immune response, response to bacterium and inflammatory processes. In addition, macrophage m2 and mast cell have been shown to be related with poor prognosis in a variety of cancer types. We found significantly increased infiltration level of macrophage m2 and mast cell in the immune high-risk group. We showed that patients with immune-effective tumors reflected poor prognosis for CRC and validated it in multiple independent data sets. Based on current findings, immune cell subsets might play a role for the prognosis differences observed between risk groups as defined by our immune signature. Furthermore, we found that the complementary value of immune signature and clinical characteristics allowed an even better accurate stratification of patients in terms of relapse-free survival in CRC.

Given the heterogeneity of colorectal cancer, patients are at substantial risk for recurrence and death, even after complete surgical resection. The management of adjuvant chemotherapy therapy in early-stage CRC remains controversial. Therefore, it is important to develop an individualized approach for the management of CRC. Reliable prognostic biomarkers can identify patients with poor outcomes irrespective of treatment, predictive biomarkers inform patients who might benefit from additional systemic therapy, and thus have more direct clinical relevance. During the last decade, significant research on gene expression-based prognostic signatures has shown a great prognostic power in CRC. Several gene expression scoring assays for recurrence risk prediction have been proposed, including Oncotype DX Colon Cancer,30 ColoPrint34 and ColoGuideEx.35 However, limited accuracy of survival estimation hampers effective clinical implementation of these methods. One common drawback of all current assays lies in normalization of expression profile given rise to technical biases inherent across different platforms with RNA-Seq or microarray. The immune-related gene pairs

### Table 1. Patient demographics and clinical characteristics for the training and validation cohorts.

|                  | Training cohort | Meta-validation cohorts |
|------------------|-----------------|-------------------------|
|                  | CIT (GSE39582, n = 565) | TCGA (n = 572) | De Sousa (GSE33113, n = 90) | Jorissen (GSE14333, n = 290) | Smith (GSE17536, n = 177) | Kirzin (GSE39084, n = 68) |
| Age, mean        | 66.85           | 66.43                   | 70.03                   | 65.95               | 65.48               | 59.71                   |
| Gender           |                 |                         |                         |                    |                    |                         |
| Female           | 256 (45%)       | 269 (47%)               | 48 (53%)                | 126 (43%)            | 81 (46%)            | 35 (51%)                |
| Male             | 309 (55%)       | 303 (53%)               | 42 (47%)                | 164 (57%)            | 96 (54%)            | 33 (49%)                |
| Location         |                 |                         |                         |                    |                    |                         |
| L                | 342 (61%)       |                         | 161 (56%)               | 36 (53%)             |                    |                         |
| R                | 223 (39%)       |                         | 128 (44%)               | 31 (46%)             |                    |                         |
| Differentiation  |                 |                         |                         |                    |                    |                         |
| Well             | 2 (2%)          |                         | 16 (9%)                 |                    |                    |                         |
| Moderately       | 60 (67%)        |                         | 134 (76%)               |                    |                    |                         |
| Poorly           | 25 (28%)        |                         | 27 (15%)                |                    |                    |                         |
| Unknown          | 3 (3%)          |                         | 3 (3%)                  |                    |                    |                         |
| Stage            |                 |                         |                         |                    |                    |                         |
| I                | 32 (6%)         | 99 (17%)                | 90 (100%)               | 94 (91%)            | 94 (14%)            | 8 (12%)                 |
| II               | 264 (47%)       | 208 (36%)               | 90 (100%)               | 94 (91%)            | 94 (14%)            | 8 (12%)                 |
| III              | 205 (36%)       | 169 (30%)               | 91 (31%)                | 91 (31%)            | 91 (31%)            | 16 (24%)                |
| IV               | 60 (11%)        | 81 (14%)                | 61 (21%)                | 61 (21%)            | 61 (21%)            | 20 (29%)                |
| Unknown          | 4 (1%)          | 15 (3%)                 |                        |                    |                    | 1 (1%)                  |
| Chemotherapy     |                 |                         |                         |                    |                    |                         |
| No               | 315 (56%)       |                         | 171 (59%)               |                    |                    |                         |
| Yes              | 233 (41%)       |                         | 118 (41%)               |                    |                    |                         |
| Unknown          | 17 (3%)         |                         |                        |                    |                    |                         |
| MMR              |                 |                         |                         |                    |                    |                         |
| MSI              | 75 (13%)        | 80 (14%)                | 25 (28%)                |                    |                    | 19 (28%)                |
| MSS              | 443 (78%)       | 478 (84%)               | 65 (72%)                |                    |                    | 49 (72%)                |
| Unknown          | 47 (8%)         | 14 (2%)                 |                        |                    |                    |                         |
| KRAS mutation    |                 |                         |                         |                    |                    |                         |
| WT               | 328 (58%)       | 239 (42%)               | 70 (78%)                |                    |                    | 40 (59%)                |
| MUT              | 216 (38%)       | 127 (22%)               | 20 (22%)                |                    |                    | 28 (41%)                |
| Unknown          | 21 (4%)         | 206 (36%)               |                        |                    |                    |                         |
| BRAF mutation    |                 |                         |                         |                    |                    |                         |
| WT               | 460 (81%)       | 329 (58%)               | 73 (81%)                |                    |                    | 60 (88%)                |
| MUT              | 51 (9%)         | 37 (6%)                 | 17 (19%)                |                    |                    | 8 (12%)                 |
| Unknown          | 54 (10%)        | 206 (36%)               |                        |                    |                    |                         |
signature is based on the relative ranking of gene expression values, which only refers to the pairwise comparison of the gene expression profile within a sample. Our prognostic signature was derived from immune-related genes, while the signature genes of previous studies were based on the whole genome level screening, therefore, there was only a low level of overlap between our signature and that of other teams. This strategy is specifically designed to perform processing combined gene expression profiles from multiple data sets robustly. CRC patients with clinically defined groups (e.g., tumor stage) can be further stratified into subgroups with different survival outcomes by our prognostic immune signature, which outperformed the commercialized molecular biomarker. For that reason, our signature can evaluate CRC prognosis in an individualized, single-sample form and be expediently promoted to clinical usage.

There are some limitations to our study. First, the prognostic signature is based on the gene expression profiles produced by RNA-seq or microarray platforms, which is difficult to popularize in daily clinical applications due to its high price, long conversion cycle and requirement of bioinformatics expertise. Second, the training data set used to establish the immune signature came from retrospective studies and included fresh frozen samples. Therefore, there are still doubts about the stability and efficiency of FFPE samples. More data sets with different sample attributes need to be included for broader validation.

Taken together, immune gene signature identified by us may provide a legitimate approach in CRC management. Meanwhile, we also revealed that the signature positively correlated with the infiltration of immune cell subsets and inflammatory response (e.g., macrophage m2, mast cell and immune response). Our immune gene signature can effectively predict patient survival of CRC patients. Moreover, this signature provides a panoramic view of the tumor immune microenvironment and will be a useful predictive tool to identify patients who might benefit from immunotherapy.

Materials and methods

Public datasets

This study used public data to do comprehensive analysis (Figure S1). The complete lists of selected all gene expression profiles (GEP), related accession numbers and corresponding publications are given in Tables S1. In total, six studies were selected, including The Cancer Genome Atlas (TCGA) cohort (n = 572),36 CIT (GSE39582, n = 565),15 De Sousa (GSE33113, n = 90),17 Jorissen (GSE14333, n = 290),37 Smith (GSE17536, n = 177) and Kirzin (GSE39084, n = 68).39 Data from The Cancer Genome Atlas (TCGA) were downloaded from the FireBrowse website (http://firebrowse.org) in March 2018. For the other cohorts, gene expression data together with clinical profiles were downloaded from Gene Expression Omnibus. CIT cohort was used as discovery and training due to its relatively high-quality clinical records and long-term follow-up. Other data sets were merged as meta-validation
cohort to verify the robustness of the IRGPs. Patients with chemotherapy treatment were excluded from the discovery data set and only patients with valid survival information were used in all studies. In total, 1,762 cases were analyzed in this study.

**Gene expression data processing**

The publicly available data sets with CRC tumor samples from the Gene Expression Omnibus were normalized using the robust multi-array average (RMA) method as implemented in the ‘affy’ package. For CIT cohort, GEP was first normalized using the RMA method and non-biological effects across batches were corrected using Combat method, using series’ identifiers as batch variables and no covariates. Level 3 TCGA RNA-seq data for colon and rectal cancer downloaded in Transcripts Per Million (TPM) processed by RSEM were further log-transformed, and non-tumor samples were removed. For each data set, the expression profiles were collapsed from probe sets to genes without further normalization. Only patients with complete survival information were used for the analyses. No further normalization methods were used for merging different data sets.

**Identification of CRC specific IRGPs for prognosis prediction**

The identification of prognostic IRGPs was performed as described previously. CIT cohort was used as the discovery data set and trained the model. We downloaded 1,811 unique immune-related genes (IRGs) from the ImmPort database (https://immport.niaid.nih.gov) accessed on 1/30/2018. In total, IRGs constitutes 17 categories including natural killer cell cytotoxicity, presentation pathways, cytokines, cytokine receptors, and antigen processing. IRGs measured by all platforms with relatively high variation (determined by MAD > 0.5, at this cutoff roughly 30% top variation genes were kept) in the discovery set were selected. Each IRGP was computed by pairwise comparison the gene expression level in a specific sample or profile. More specifically, in a pairwise comparison, the output is 1 if the first element is larger than the later one and 0 for the different order. After removing IRGPs with a relatively small variation (MAD = 0, which means almost no changes across patients), the remaining IRGPs were left and selected as initial candidate IRGPs for prognosis prediction.

**Construction of IRGPI for prognosis prediction**

The construction of prognostic signature was performed as described previously. From the initial candidate IRGPs, an IRGP index (IRGPI) was constructed using Lasso Cox proportional hazards regression and 19 gene pairs were used to define the final model. To stratify patients into low or high-risk groups, the optimal cut off of the IRGPI was determined using time-dependent receiver operating characteristic (ROC)
curve analysis (survivalROC, version 1.0.3) at 5 years in the training cohort for relapse-free survival (RFS).

Validation of the IRGPI

The IRGPI prognostic value was evaluated in all stages CRC patients and early stages group (stage I and II) in the training, meta-validation (combination of the TCGA, Jorissen, and De Sousa, Smith and Kirzin cohorts) and independent validation cohorts by the log-rank test. We then combined IRGPI with other clinical factors in the multivariate analysis. Age and stage were treated as continuous variables. Stage I was transformed into 1; Stage II was transformed into 2; Stage III was transformed into 3 and Stage IV was transformed into 4. The prognostic accuracy of IRGPI was estimated using the concordance index (C-index), which range from 0 to 1.0. The prognostic efficiency of IRGPI was compared with the existing multigene signature Oncotype Dx Colon Cancer with C-index.

Profiling of infiltrating immune cells

To dissect immune cell infiltration in different risk groups, we employed CIBERSORT, a popular algorithm for characterizing cell composition from bulk-tumor gene expression profiles. Based on a set of reference gene expression values, CIBERSORT deconvolutes bulk tumor samples with a minimal representation for each cell type using support vector regression. More specifically, normalized gene expression data sets were uploaded to the CIBERSORT web portal (http://cibersort.stanford.edu/) for analysis, using the default signature matrix at 1,000 permutations. For each sample, CIBERSORT inferred the relative proportions of 22 types of infiltrating immune cells consisting T cells, B cells, natural killer cells, macrophages, plasma cells, neutrophils, dendritic cells, and eosinophils, amongst others.

Gene ontology (GO) analysis

Gene ontology (GO) analysis was performed in gProfiler for the prognostic immune signature. FDR-adjusted $P < 0.05$ was used to select statistically significant gene sets.

Construction and validation of a composite immune-clinical prognostic index

We integrated age, stage, sex and IRGPI risk score to the immune-clinical prognostic index (ICPI) using Cox proportional hazards regression in training cohort as following: ICPI = (1.085 × sex) + (0.737 × stage) + (2.464 × IRGPI score). The prognostic efficiency of ICPI was compared with the C-index of IRGPI and depicted by the restricted mean survival (RMS) curve. High RMS time ratio represents large prognostic difference.
Statistical analysis

Statistical analyses were performed using R (version 3.4.3, www.r-project.org). Continuous variables were expressed as mean standard error of the mean and were compared using student’s t-tests or Wilcoxon rank sum tests. Survival analyses were performed using the Kaplan–Meier method and compared using a log-rank test by the ‘survival’ package (version: 2.41.3). The RMS curve and time ratio were also calculated by survival package. Hazard ratios and C-index were calculated using survcomp package (version: 1.28.4). For all tests, a p-value < 0.05 was considered to be significant. Statistical significance is shown as *p < 0.05, **p < 0.01, ***p < 0.001.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81572307) and Major Project of Medical and Health Technology Development Program in Zhejiang Province (No. 7211902).

References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013;49(6):1374–1403. doi:10.1016/j.ejca.2012.12.027.
2. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin. 2014;64(2):104–117. doi:10.3322/caac.21220.
3. Verma M, Kumar V. Epigenetic biomarkers in colorectal cancer. Mol Diagn Ther. 2016;21(2):153–165. doi:10.1007/s40291-016-0244-x.
4. Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. Genet Med. 2017;19(3):314–321. doi:10.1038/gim.2016.101.
5. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, Ostos LCG, Lannon WA, Grotzinger C, Del Rio M, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013;19(5):619–625. doi:10.1038/nm.3175.
6. Inadomi JM. Screening for colorectal neoplasia. N Engl J Med. 2017;376(2):149–156. doi:10.1056/NEJCm1512286.
7. Chang W, Gao X, Han Y, Du Y, Liu Q, Wang L, Tan X, Zhang Q, Liu Y, Zhu Y, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. Gut. 2014;63(9):1457–1467. doi:10.1136/gutjnl-2013-305475.
8. Lamprecht S, Kaller M, Schmidt EM, Blaj C, Schiergens TS, Engel J, Jung A, Hermeking H, Grünewald TGP, Kirchner T, et al. PBX3 is part of an EMT regulatory network and indicates poor outcome in colorectal cancer. Clin Cancer Res. 2018 Feb 1;24:1974–1986. doi:10.1158/1078-0432.CCR-17-2572.
9. Taieb J, Kourie HR, Emile J-F, Le Malicot K, Balogoun R, Taberneró J, Mini E, Polprecht G, Van Laethem J-L, Mulot C, et al. Association of prognostic value of primary tumor location in stage III colon cancer with RAS and BRAF mutational status. JAMA Oncol. 2017 Nov 22. doi:10.1001/jamaoncol.2017.3695.

Figure 4. Kaplan–Meier curve of RFS and restricted mean survival (RMS) curves for patients stratified by the IRGPI and the ICPI. CRC patients were ranked by immune risk scores in the CIT cohort (a) and the meta-validation cohorts (d). Patients were stratified with ICPI in all patients of CIT (b) and all patients of five independent validation datasets (e). The RMS curve for IRGPI and ICPI scores was plotted for the CIT (c) and all patients of five independent validation datasets (f).
10. Barrier A, Boelle P-Y, Roser F, Gregg J, Tse C, Brault D, Lacaine F, Houry S, Hugueri M, Franc B, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol. 2006; 24(29):4685–4691. doi:10.1200/JCO.2005.05.0229.

11. Wang Y, Jatkoe T, Zhang Y, Mutch MG, Talantov D, Jiang J, McLeod HL, Atkins D. Gene expression profiles and molecular markers to predict recurrence of Duke’s B colon cancer. J Clin Oncol. 2004;22(9):1564–1571. doi:10.1200/JCO.2004.08.0400.

12. Yothers G, O’Connell MJ, Lee M, Lopatin M, Clark-Langone KM, Millward C, Paik S, Sharif S, Shak S, Wolmark N. Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and leucovorin (FU/LV) and FU/LV plus oxaliplatin. J Clin Oncol. 2013;31(36):4512–4519. doi:10.1200/JCO.2013.49.0219.

13. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, Lopez-Doriga A, Santos C, Marijnen C, Westera J, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol. 2011;29(17):17–24. doi:10.1200/JCO.2010.30.1077.

14. Subramanian J, Simon R. Response: re: gene expression-based prognostic signatures in lung cancer: ready for clinical use? JNCI J Natl Cancer Inst. 2013;105(5):476–478. doi:10.1093/jnci/djt386.

15. Marisa L, de Reyniés A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi M-C, Schiappa R, Guenot D, Ayadi M, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10(5):e1001453. doi:10.1371/journal.pmed.1001453.

16. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, Kerr D, Aaltenon LA, Arango D, Kruhoffer M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage B and C colorectal cancer. Clin Cancer Res. 2009;15(24):7642–7651. doi:10.1158/1078-0432.CCR-09-0547.

17. de Sousa E, Melo F, Colak S, Buikhuisen J, Koster J, Cameron K, de Jong HJ, Tuynman JB, Prasetyanti PR, Fessler E, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. Cell Stem Cell. 2011;9(5):476–485. doi:10.1016/j.stem.2011.07.010.

18. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Genet. 2010;11(10):733–739. doi:10.1038/ng.2825.

19. Popovici V, Budinska E, Tejpár S, Weinrich S, Estrella H, Sinkkonen L, Zhou JX, Kreisberg R, Kauffman SA, Huang S, et al. Somatic copy number alterations in colorectal cancers reveal distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol. 2015;16:64. doi:10.1186/s13059-015-0667-4.

20. O’Connell MJ, Lavery I, Yothers G, Paik S, Clark-Langone KM, Lopatin M, Watson G, Baehner FL, Shak S, Baker J, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol. 2010;28(25):3937–3944. doi:10.1200/JCO.2010.29.9538.

21. Niño D, Komohara Y, Murayama T, Aoki R, Kimura Y, Hashikawa K, Kiyasu J, Takeuchi M, Suefuji N, Sugita Y, et al. Ratio of M2 macrophage expression is closely associated with poor prognosis for Angioimmunoblastic T-cell lymphoma (AITL). Pathol Int. 2010;60(4):278–283.

22. Malletton A, Silvestris N, Saponaro C, Ranieri G, Russo A, Caruso S, Popescu O, Simone G, Paradiso A, Mangia A. High density of tryptase-positive mast cells in human colorectal cancer: a prognostic factor related to prostate-activated receptor 2 expression. J Cell Mol Med. 2013;17(8):1025–1037. doi:10.1111/jcmm.2013.17.issue-8.

23. Goasguen JE, Zandeeni M, Mathiott C, Scheiff HM, Bize W, Ly-Sunnarum B, Grosbois M, Moncouturier M, Michaux J, et al. Mature plasma cells as indicator of better prognosis in multiple myeloma. New methodology for the assessment of plasma cell morphology. Leuk Res. 1999;23:1133–1140.

24. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, Lopez-Doriga A, Santos C, Marijnen C, Westera J, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol. 2011;29(17):17–24. doi:10.1200/JCO.2010.30.1077.

25. Agenes TH, Sveen A, Merok MA, Lind GE, Nesbakken A, Skothem RI, Lothe RA. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. Gut. 2012;61(11):1560–1567. doi:10.1136/gutjnl-2011-301179.

26. The cancer genome atlas network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330–337. doi:10.1038/nature11252.

27. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, Kerr D, Aaltenon LA, Arango D, Kruhoffer M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage B and C colorectal cancer. Clin Cancer Res. 2009;15(24):7642–7651. doi:10.1158/1078-0432.CCR-09-0547.

28. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, Lu P, Johnson JC, Schmidt C, Bailey CE, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010;138(3):958–968. doi:10.1053/j.gastro.2009.08.053.

29. Kirzin S, Marisa L, Guimbaud R, De Reynies A, Legrain M, Laurent-Puig P, Cordelier P, Pradère B, Bonnet D, Meggetto F, et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. PLoS One. 2014;9(8):e103159. doi:10.1371/journal.pone.0103159.
40. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of affymetrix genechip data at the probe level. Bioinformatics. 2004;20(3):307–315. doi:10.1093/bioinformatics/btg405.

41. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–127. doi:10.1093/biostatistics/kxj037.

42. Bhattacharya S, Andor S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res. 2014;58(2–3):234–239. doi:10.1007/s12026-014-8516-1.

43. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56(2):337–344. doi:10.1111/j.0006-341X.2000.00337.x.

44. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–457. doi:10.1038/nmeth.3337.

45. Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. Oncotarget. 2015;6(34):36308–36318. doi:10.18632/oncotarget.v6i34.

46. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27(22):3206–3208. doi:10.1093/bioinformatics/btr511.