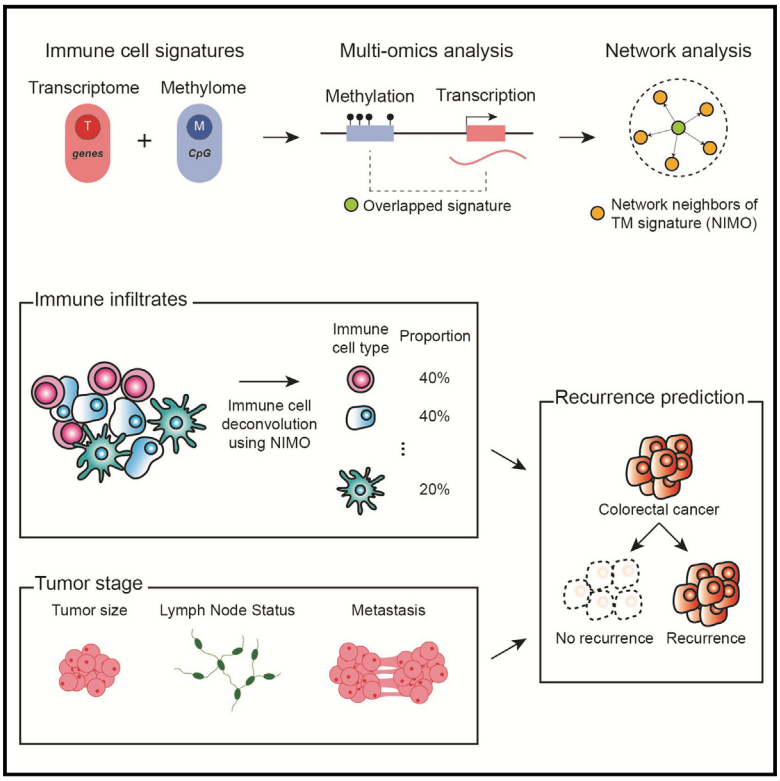
Article

Patterns

Information about immune cell proportions and tumor stage improves the prediction of recurrence in patients with colorectal cancer

Graphical abstract



Highlights

- Combining tumor stage and immune cell proportions improves CRC recurrence prediction
- Validation of the recurrence prediction was performed using two independent cohorts
- Network analysis of multiomics data was used to identify robust immune cell markers

Authors

JungHo Kong, Jinho Kim, Donghyo Kim, ..., Yong Beom Cho, Donghyun Park, Sanguk Kim

Correspondence

gscyb@skku.edu (Y.B.C.), eastwise37@gmail.com (D.P.), sukim@postech.ac.kr (S.K.)

In brief

Accurately predicting CRC recurrence is crucial for treatment decisions, but current methods have limited success. There is an urgent need to identify additional predictive features. Combining immune cell proportions and TNM stage information improved recurrence prediction. Specifically, we identified robust immune cell biomarkers using transcriptome and methylation data. These findings highlight the potential clinical utility of immune cell profiling in cancer prognosis.

Kong et al., 2023, Patterns *4*, 100736
June 9, 2023
https://doi.org/10.1016/j.patter.2023.100736

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Information about immune cell proportions and tumor stage improves the prediction of recurrence in patients with colorectal cancer

JungHo Kong,[1,9] Jinho Kim,[2,9] Donghyo Kim,[1,9] Kwanghwan Lee,[1,9] Juhun Lee,[1,9] Seong Kyu Han,[1,9] Inhae Kim,[6,7] Seongsu Lim,[8] Minhyuk Park,[1] Seungho Shin,[4] Woo Yong Lee,[5] Seong Hyeon Yun,[5] Hee Cheol Kim,[5] Hye Kyung Hong,[6] Yong Beom Cho,[5,7,*] Donghyun Park,[4,*] and Sanguk Kim[1,3,8,10,*]

[1]Department of Life Sciences, Pohang University of Science and Technology, Pohang 37673, Korea
[2]Precision Medicine Center, Future Innovation Research Division, Seoul National University Bundang Hospital, Seongnam 13620, Korea
[3]Institute of Convergence Science, Yonsei University, Seoul 120-749, Korea
[4]GENINUS, Inc., Seoul 05836, Korea
[5]Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea
[6]Institute for Future Medicine, Samsung Medical Center, Seoul 06351, Korea
[7]Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul 06351, Korea
[8]School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang 37673, Korea
[9]These authors contributed equally
[10]Lead contact
*Correspondence: gscyb@skku.edu (Y.B.C.), eastwise37@gmail.com (D.P.), sukim@postech.ac.kr (S.K.)
https://doi.org/10.1016/j.patter.2023.100736

THE BIGGER PICTURE Colorectal cancer is a significant global public health issue, and accurately predicting its recurrence remains a challenge despite advances in screening and treatment. Accurate recurrence prediction is crucial for clinicians to make informed decisions about treatment and follow-up care, leading to timely interventions that may improve outcomes and potentially prolong survival. In this study, we present a method that combines immune cell information with clinical data to enhance the accuracy of recurrence risk prediction. The predictive performance of the method was validated using two independent cohorts of patients with colorectal cancer of different races. Our approach has implications for both clinical practice and research, as it helps to suggest treatment strategies by predicting recurrence risk and identifying potential therapeutic targets based on immune cell information.

Development/Pre-production: Data science output has been rolled out/validated across multiple domains/problems

SUMMARY

Predicting cancer recurrence is essential to improving the clinical outcomes of patients with colorectal cancer (CRC). Although tumor stage information has been used as a guideline to predict CRC recurrence, patients with the same stage show different clinical outcomes. Therefore, there is a need to develop a method to identify additional features for CRC recurrence prediction. Here, we developed a network-integrated multiomics (NIMO) approach to select appropriate transcriptome signatures for better CRC recurrence prediction by comparing the methylation signatures of immune cells. We validated the performance of the CRC recurrence prediction based on two independent retrospective cohorts of 114 and 110 patients. Moreover, to confirm that the prediction was improved, we used both NIMO-based immune cell proportions and TNM (tumor, node, metastasis) stage data. This work demonstrates the importance of (1) using both immune cell composition and TNM stage data and (2) identifying robust immune cell marker genes to improve CRC recurrence prediction.

INTRODUCTION

A major challenge in cancer treatment is the prediction of cancer recurrence. This is especially important for colorectal cancer (CRC), since CRC is the second leading cause of cancer-associated deaths in the USA.[1] Traditional tumor staging systems, which consider tumor burden (T), cancer cells in lymph nodes (N), and distant metastasis (M), have been the most reliable and frequently used strategies to predict recurrence in patients with CRC.[2] However, the clinical outcomes of patients with CRC with the same TNM stage vary significantly,[2–7] suggesting that additional parameters may be required to improve the prediction of recurrence for patients with CRC.

An emerging feature for predicting cancer recurrence is the tumor microenvironment.[8] Indeed, infiltration of immune cells was found to be a strong predictor of clinical outcome-related factors, such as prognosis and cancer recurrence in patients with CRC.[2,4–7,9–11] For example, dense infiltration of memory T cells is correlated with a lower incidence of tumor recurrence in patients with CRC,[7] suggesting that accurate estimation of infiltrated immune cell levels can be leveraged for clinical usage. Accordingly, numerous methods have been developed that aim to detect immune infiltrates, ranging from conventional immunostaining-based cell counting methods to omics-based immune cell deconvolution methods.[12,13] Detection of immune cell infiltrates depends on lists of marker genes (immune cell signature genes)[14–16]; therefore, carefully selected immune signature genes are needed to predict recurrence in patients with CRC.

Integrating multiple omics layers provides a unique opportunity to help detect immune cell signature genes. Previous studies found that multiomics factors are predictive of immune-related adverse events caused by immune checkpoint inhibitors.[17] Moreover, Webber et al. found that a multiomics gene module detected by integrating protein, mutation, copy number, methylation, and expression-level data was highly predictive of drug response.[18] Similarly, a multiomics-based neural network model was more predictive of anticancer treatment responses than single omics-based models,[19] suggesting the importance of utilizing multiple omics layers for identifying important gene markers. Altogether, the integration of multiomics data can help detect robust signals, although it is still unclear whether incorporating various omics layers can improve immune cell deconvolution.

Along with a multiomics approach, network-based methods can be applied to successfully identify immune cell signature genes. We recently reported that protein-protein interaction (PPI) networks can be leveraged to identify biomarkers associated with drug sensitivity from the pharmacogenomics data of patient-derived cancer organoids.[20] Because previous studies have shown that the genes that are proximal to each other in a PPI network are associated with similar phenotypes,[21,22] it is highly likely that signature genes associated with a cell type would also form clusters in a PPI network. However, a method that utilizes both multiomics data and a network-based approach to identify immune cell signature genes has not been explored.

In this study, we present a network-integrated multiomics (NIMO) approach to identify immune cell marker genes for immune cell deconvolution. Using NIMO marker genes, we predicted cancer recurrence using the NIMO approach in two retrospective studies of 114 and 110 patients with CRC. We found statistically significant improvement in prediction when using both TNM stage and immune infiltrate composition data inferred from the NIMO-based approach compared with prediction using only TNM stage data. Specifically, prediction improvements were observed in terms of accuracy, precision, recall, and F1 score, which suggests that improvements were made by increased prediction of true positives and true negatives. Among the 22 immune cell types considered, we found that high proportions of regulatory T (T_{reg}) cells and M2 macrophages were associated with recurrence in our CRC dataset. We also observed the cooccurrence of T_{reg} cells and M2 macrophages and mutually exclusive patterns between T_{reg} cells and M1 macrophages and T_{reg} cells and CD8 T cells; these results suggest that our immune gene signature accurately identified known crosstalk between immune cells. Additionally, we found that the expression level of NIMO-based T_{reg} cell markers correlated with that of FOXP3, a known marker of T_{reg} cells, in external CRC single-cell RNA sequencing data, supporting the idea that our procedure for the selection of immune cell signature genes can identify robust immune cell markers. Altogether, our results indicate the importance of (1) understanding the immune cell composition and (2) using an appropriate procedure to select immune cell signature genes to improve the prediction and understanding of recurrence in patients with CRC.

RESULTS

NIMO-based immune cell composition information combined with TNM stage data improves the prediction of recurrence in patients with CRC

We curated a retrospective cohort of 114 patients with CRC (Samsung Medical Center [SMC] cohort), composed of 47 patients with cancer recurrence and 67 patients without recurrence (Figure S1). To identify the immune cell composition in patients with CRC using bulk RNA sequencing data, we first detected immune cell-specific marker genes. We designed a custom immune cell signature gene detection method that utilizes a NIMO approach (Figure 1A). Briefly, we first filtered transcriptome-based immune cell signatures by comparing the methylation-based immune cell signatures. Among the transcriptome-based signature genes from Newman et al. (LM22),[12] we filtered genes whose proximal CpG sites were included in the signature CpGs of Chakravarthy et al.[13] (see experimental procedures). Additionally, we further included differentially expressed genes (DEGs) that were first neighbors in the PPI network to the genes selected in the previous procedure. The selected genes for each immune cell type are provided in Figure S2. The selected genes were then used to determine immune cell proportions in patients with CRC.

To validate the gene sets generated by the NIMO approach, we examined whether the identified signatures covered markers of various immune cell types. We conducted enrichment tests of TM signatures with and without network expansion on known cell-type markers using the following two databases: CellMarker[23] and PanglaoDB.[24] Our findings indicate that TM signatures with network expansion were significantly enriched in markers of all cell types (Figure S3; $p < 0.05$, hypergeometric test), whereas TM signatures without network expansion did not show significant enrichment. For instance, TM signatures with

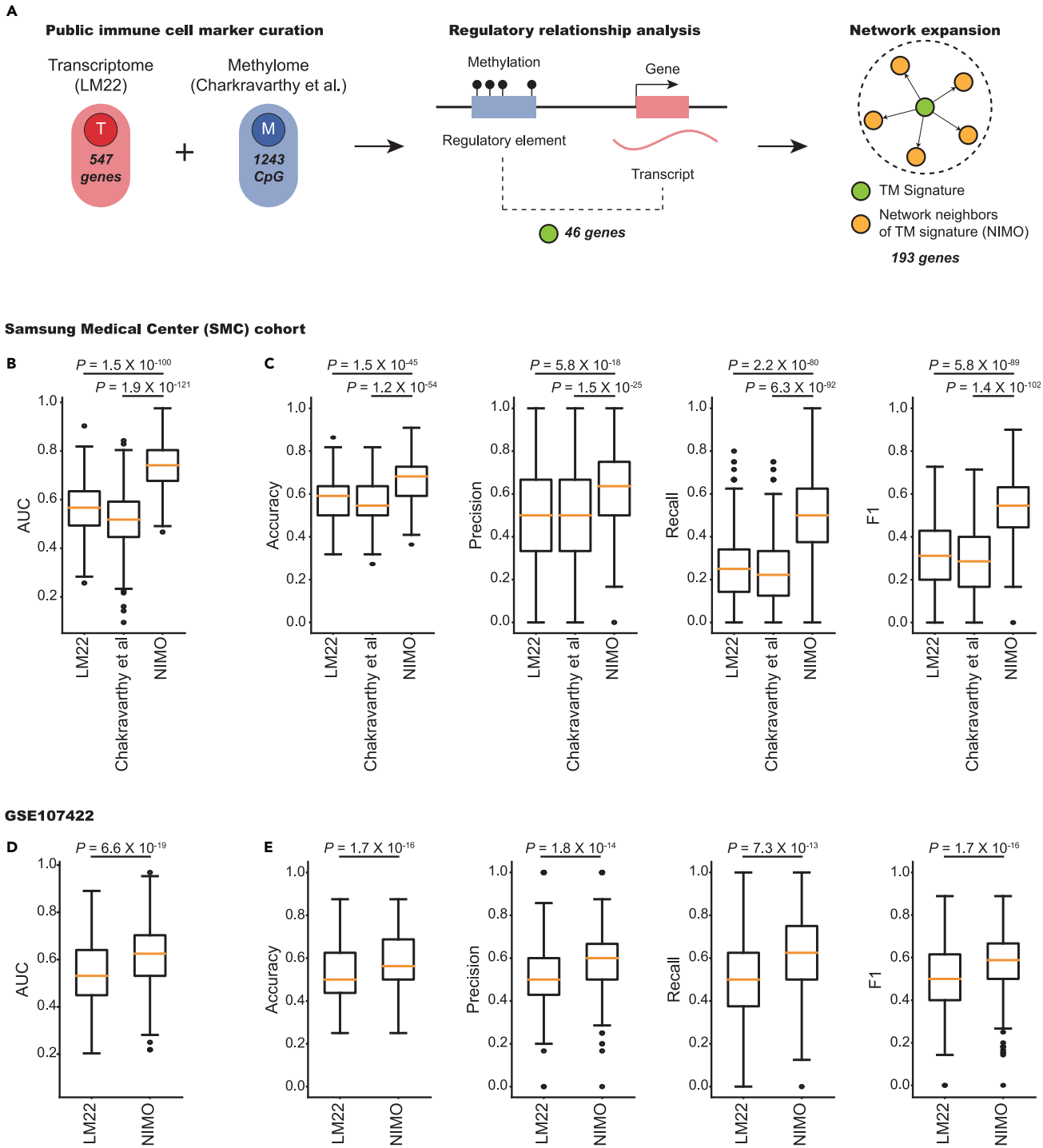


Figure 1. Overview of the NIMO approach and the ability of the NIMO approach and other immune cell signatures to predict CRC recurrence

(A) Immune cell markers were curated from Newman et al. (LM22) and Charkravarthy et al. for transcriptome (T)- and methylome (M)-based signatures, respectively. The T and M signature genes that were determined to have regulatory relationships were further identified (TM signature genes; 46 genes). The first neighbors of the TM signature genes that were included in the T signature were chosen as NIMO signature genes.

(B and C) CRC recurrence in the SMC cohort (recurrence: 47, nonrecurrence: 67) was predicted using the LM22, Charkravarthy et al., or NIMO signatures for immune cell deconvolution.

(D and E) CRC recurrence of the GEO: GSE107422 cohort (recurrence: 38, nonrecurrence: 72) was predicted using LM22 or NIMO signatures for immune cell deconvolution. Deconvolution using the M-based signature, Charkravarthy et al., was not available since methylation data for the GEO: GSE107422 cohort were not reported. Statistical differences in prediction performances were measured by the Mann-Whitney U test.

See also Figures S1–S7.

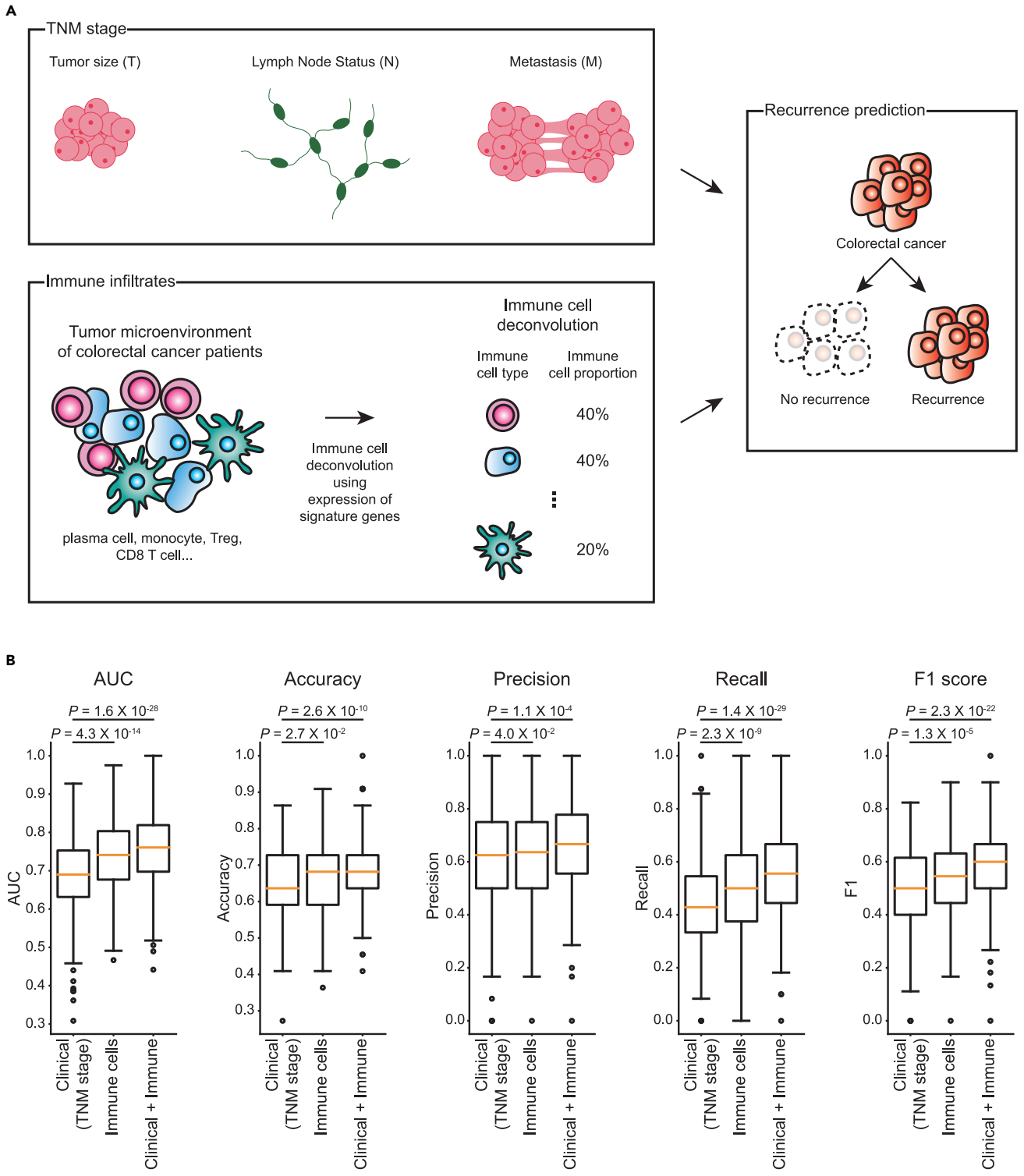


Figure 2. Prediction of CRC recurrence using immune cell proportions inferred from NIMO signature gene and TNM stage data

(A) Prediction of CRC recurrence from estimated immune cell proportions. Immune cell proportions were estimated using the immune cell gene signature inferred from the NIMO-based approach as an input for cell-type deconvolution. The following 22 target immune cell types were used in the analysis: B cells (naive, memory, and plasma cells); monocytes; macrophages (M0, M1, and M2); T cells (CD4 memory resting, CD4 memory activated, CD4 naive, Treg, follicular helper, CD8, and γδ); NK cells (resting and activated); mast cells (resting and activated); dendritic cells (resting and activated); neutrophils; and eosinophils.

(legend continued on next page)

network expansion effectively covered macrophage marker genes annotated in PanglaoDB (17 genes, $p = 6.3 \times 10^{-11}$), while those without expansion did not exhibit significant enrichment (1 gene, $p = 0.84$).

We also examined whether NIMO signature gene expression was significantly correlated with DNA methylation (Figure S4). Our analysis revealed that the gene sets selected by the NIMO approach (TM signature) contained significantly more genes with a negative correlation between expression and methylation than the filtered gene set (T-TM signature) ($p = 2.6 \times 10^{-3}$, Fisher's exact test). These findings suggest that the NIMO approach effectively filtered out transcripts regulated by DNA methylation to a certain extent, although not completely.

We found that the selection of immune cell signature genes had a significant impact on the ability to predict CRC recurrence. To evaluate the prediction performance, we used transcriptome- and methylome-based immune cell markers from Newman et al. (LM22)[12] and Chakravarthy et al.,[13] respectively, to estimate immune cell proportions. We conducted cross-validation using four different machine learning models (support vector machine, extra trees, random forest, and logistic regression models) and an ensemble model that integrated prediction results into a single classifier. We evaluated prediction performance using measures such as the area under the receiver operating curves (AUC), accuracy, precision, recall, and F1 score. We observed that the ensemble model outperformed all the individual machine learning (ML) models (Figure S5) and selected it for further analyses.

We found that neither the LM22 markers nor the Chakravarthy et al. markers were strong predictors of CRC recurrence. The average AUC using the NIMO markers was 0.74, whereas those using the LM22 and Chakravarthy et al. markers were 0.56 and 0.52, respectively (Figure 1B). This suggests that the selection of immune cell signature genes is critical for accurately estimating immune cell proportions and predicting cancer recurrence. The more robust prediction performance of NIMO compared with the transcriptome- and methylome-based marker sets was consistent across other performance measures (Figures 1C and S6). We also confirmed that the NIMO markers outperformed transcriptome- and methylome-based markers even when recurrence prediction was performed by classifying patients with CRC by stage (Figure S7). In summary, our results highlight the importance of selecting appropriate immune cell marker genes for the accurate prediction of CRC recurrence.

We validated the performance of NIMO markers to predict cancer recurrence using an independent cohort of 110 patients with CRC (GSE107422),[25] composed of 38 patients with recurrence and 72 patients without recurrence. We found that immune cell proportions inferred by the NIMO markers predicted CRC recurrence better than those based on LM22 markers. Specifically, we enumerated immune cell proportions using immune cell markers derived from NIMO and the transcriptome-based approach (LM22). Then, we predicted CRC recurrence using the cell proportions. The AUCs of NIMO (0.62) were significantly higher than those of LM22 (0.55) (Figure 1D; Mann-Whitney U test, $p = 6.6 \times 10^{-19}$). The accuracy, precision, recall, F1 score, and disease-free survival consistently showed the stronger performance of NIMO (Figures 1E and S8). These results suggest that NIMO is a robust approach to predicting CRC recurrence across independent cohorts.

Next, we investigated whether including immune cell composition could enhance the predictive power of TNM stage data. Our analysis revealed a significant improvement in prediction performance when both TNM stage and immune cell proportions were used compared with the predictions relying on TNM stage data alone (Figure 2A). This improvement was consistently observed across various performance metrics, including AUC, accuracy, precision, recall, and F1 score (Figure 2B; Mann-Whitney U test, $p < 0.05$ considered significant). We also compared the performance improvement in our model using immune cell composition with that using the expression of all genes. Interestingly, we found that the model incorporating immune cell composition inferred from gene expression outperformed the model using the expression of all genes (Figure S9). These findings strongly suggest that considering immune cell composition, in addition to TNM stage data, is crucial for accurately predicting CRC recurrence.

Association of inferred immune cell proportions and clinical features with CRC recurrence

Motivated by the ability to predict CRC recurrence, we investigated the associations between inferred immune cell proportions and clinical phenotypes of CRC. We found that a high proportion of T_{reg} cells and a high proportion of T follicular helper cells were the most significantly associated with recurrence and no recurrence, respectively, in CRC (Figure 3). We also found that macrophages were significantly associated with CRC recurrence (Mann-Whitney U test, $p < 0.05$). Accordingly, we observed that the abundance of M1 macrophages was significantly higher in patients with CRC without recurrence, whereas there was a significantly higher amount of M2 macrophages detected in the CRC recurrence patient group (Figure 3). Consistent with our results, high proportions of M1 and M2 macrophages were previously reported to be associated with favorable and poor prognosis in CRC, respectively.[26] We found that when using signatures from LM22, the significance of differences in T_{reg} cell and M1 macrophage fractions between patients with and without recurrence was reduced (Figure S10), suggesting that important cues of the tumor microenvironment may be lost when using less optimal signature genes for deconvolution. We also observed significantly higher proportions of resting mast cells and monocytes in patients with recurrent CRC and higher proportions of follicular helper T cells and activated CD4 memory T cells in patients with CRC without recurrence. Finally, we observed a higher proportion of CD8 T cells in patients with CRC without recurrence, although the difference failed to meet statistical significance.

Given the highly heterogeneous nature of patients with CRC,[27–31] we further aimed to subclassify patients with CRC

(B) CRC recurrence prediction was measured in terms of the following five different metrics for evaluating predictive performance: AUC, accuracy, precision, recall, and F1 score. CRC recurrence was predicted by using TNM stage data and/or immune cell proportions. Statistical differences in prediction performances were measured by the Mann-Whitney U test.
See also Figures S8 and S9.

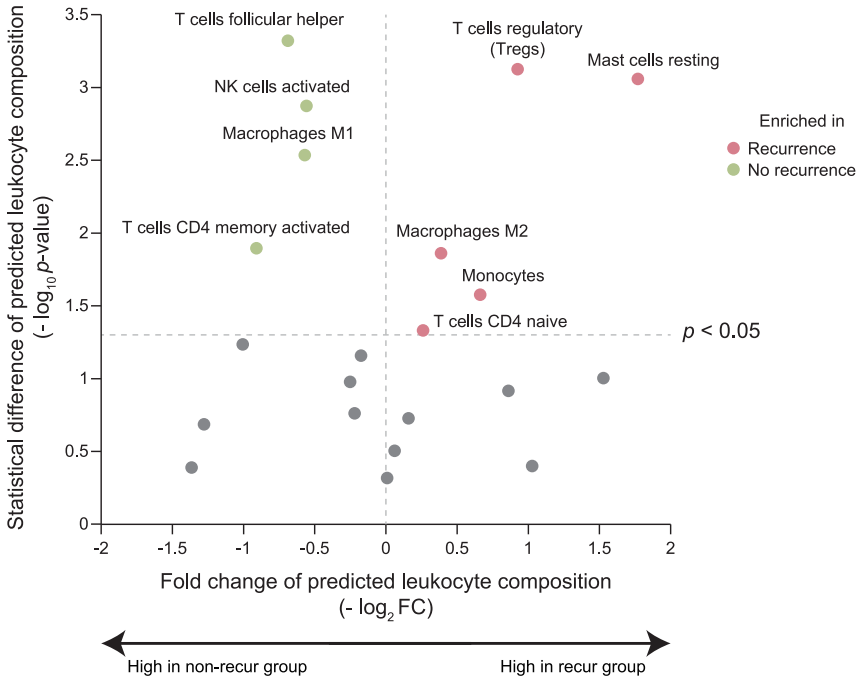


Figure 3. Association of immune cell proportions with CRC recurrence

Statistical differences in immune cell proportions between patients with CRC with and without recurrence were measured by the Mann-Whitney U test. Red and green dots indicate higher percentages of leukocytes in patients with CRC with and without recurrence, respectively. p values less than 0.05 were considered significant. Immune cell compositions were inferred using the NIMO signatures.

with and without recurrence into homogeneous subgroups using both clinical features and NIMO signature-based immune cell proportions. Using hierarchical clustering (see experimental procedures), we were able to detect three homogeneous subgroups within patients with CRC with recurrence and without recurrence (Figure 4A). Within the CRC subgroups, we found that our inferred immune cell proportions were in line with the expression of known immune cell markers. Specifically, we found that the N1 subgroup had significantly higher CD8 T cells than the group of patients with recurrence (Figure 4B; Mann-Whitney U test, $p = 3.9 \times 10^{-5}$). When testing for differential gene expression, we observed that known CD8 T cell markers,[32] including GZMB, GZMA, CD8A, and PRF1, were all significantly overexpressed in the N1 group (Figure 4B). Moreover, our analysis indicated that M1 macrophages in the R2 group were depleted compared with those in the groups of patients without recurrence (Figure 4C; $p = 1.8 \times 10^{-3}$). Reassuringly, we found that known M1 macrophage markers[33] were downregulated in the R2 group compared with the group without recurrence (Figure 4C), suggesting that the inferred immune cell proportions accurately reflect the true immune cell compositions within a patient.

Immune cell signatures inferred from the NIMO approach correspond with known biological characteristics of T_{reg} cells

As T_{reg} cells were most highly associated with recurrence in patients with CRC in our dataset, we further investigated the relationship between our signatures developed with the NIMO approach and the biological characteristics of T_{reg} cells. First, we examined whether the cooccurrence of T_{reg} cells and other immune cells was recapitulated when the proportions of immune cells were predicted using the NIMO signatures in our CRC dataset (Figure 5A). We found that cooccurrence and mutually exclusive patterns were observed between T_{reg} cells and immunosuppressive and immune-activating cells,[34] respectively, when using the NIMO signatures (Figures 5B–5D). Specifically, we identified that the predicted proportions of immunosuppressive cell types, i.e., M2 macrophages and T_{reg} cells, showed a high correlation in our patients with CRC (Mann-Whitney U test, $p = 0.046$; Figure 5B). In contrast, positive crosstalk between T_{reg} cells and M2 macrophages was not observed when considering immune cell compositions computed using gene signatures from LM22 ($p = 0.50$; Figure 5B). Similarly, we discovered statistically significant mutually exclusive patterns between T_{reg} cells and M1 macrophages ($p = 0.027$; Figure 5C) and also between T_{reg} cells and CD8 T cells ($p = 0.013$; Figure 5D) when using our signatures. Meanwhile, mutually exclusive patterns with T_{reg} cells were not observed when using immune cell signatures from LM22 for either M1 macrophages ($p = 0.13$; Figure 5C) or CD8 T cells ($p = 0.16$; Figure 5D). Taken together, our results suggest that appropriate immune cell marker selection is required to determine immune cell cooccurrence and mutually exclusive patterns.

We further validated our signatures by comparing the expression levels of a well-established T_{reg} marker gene (FOXP3) and T_{reg} signature genes defined by the NIMO approach in an external single-cell RNA sequencing dataset of patients with CRC. We observed that the average expression level of T_{reg} cell genes in our signature was moderately correlated with FOXP3 expression (Figures 5E and 5F; Spearman correlation coefficient = 0.50, $p = 5.2 \times 10^{-95}$). Our results are in line with previous reports showing that FOXP3 is the most specific marker for T_{reg} cells.[34] We also found that the significance of the correlation between FOXP3 and signature gene expression was greater for T_{reg} cell genes from the NIMO signatures than for T_{reg} cell signature genes from LM22 (Figure 5G). Taken together, our results suggest that the selection of immune cell signature genes can have an impact on immune cell deconvolution.

DISCUSSION

In this study, we found that information about the immune context can improve the prediction of CRC recurrence compared with that using only TNM stage information. Importantly, our deconvolution results revealed heterogeneous subgroups of patients with CRC, and we further validated our approach with external

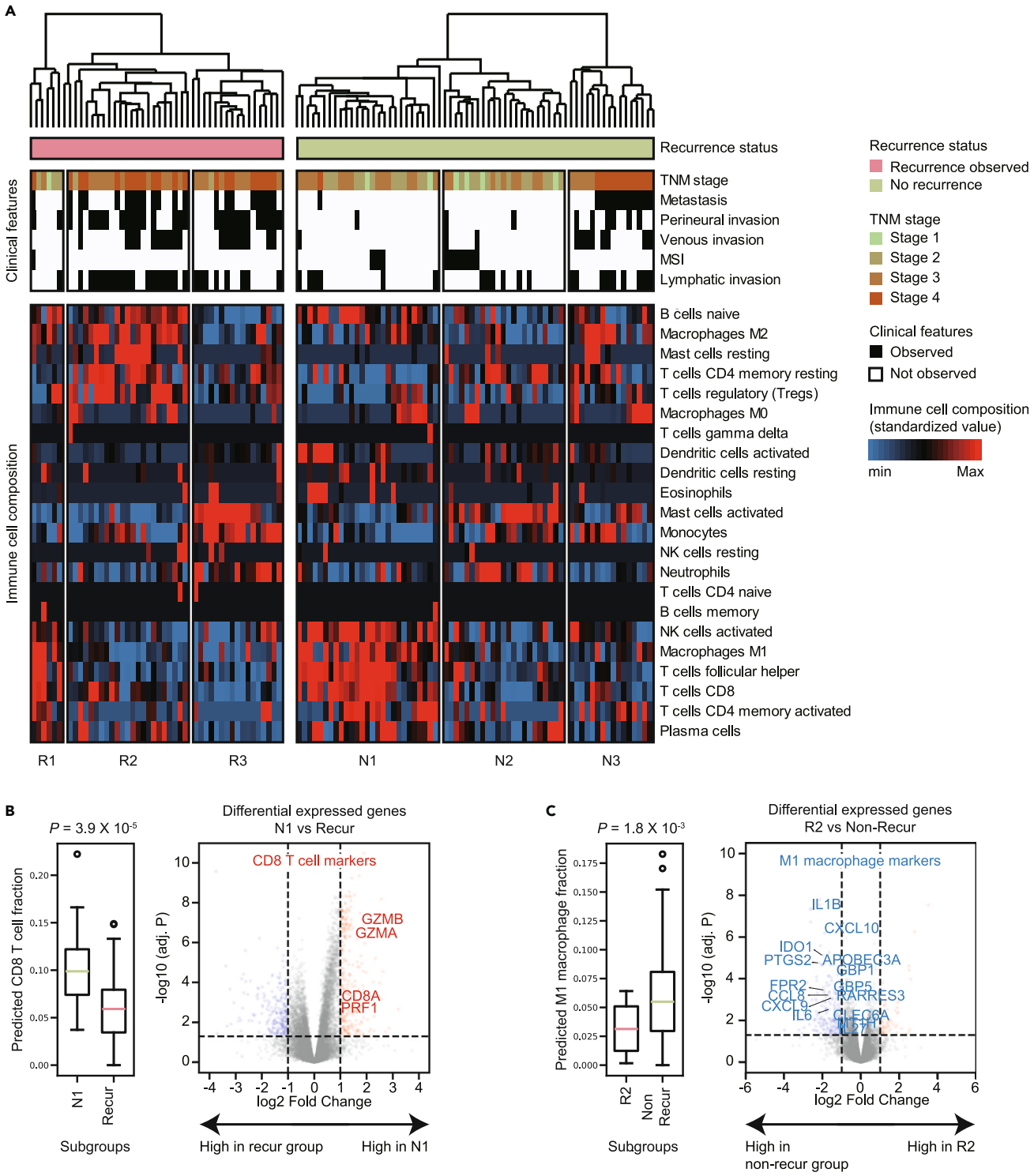


Figure 4. Subclustering of patients with CRC with and without recurrence

(A) Patients with CRC with and without recurrence were clustered based on clinical features and immune cell proportions.

(B and C) Immune cell fractions were compared between subgroups, and gene expression was compared between (B) the N1 and recurrence groups and (C) the R2 and nonrecurrence groups. For the boxplots, the Mann-Whitney U test was used to determine the significance of differences. For scatterplots, an absolute log2 fold change over 1 and an adjusted p value of under 0.05 were used to identify differentially expressed genes. Labeled genes in the scatterplots are gene markers for CD8 T cells and M1 macrophages from Jiang et al. and Beyer et al., respectively. Differentially expressed genes among the gene markers from Jiang et al. and Beyer et al. are displayed in the scatterplots.

See also Figure S10.

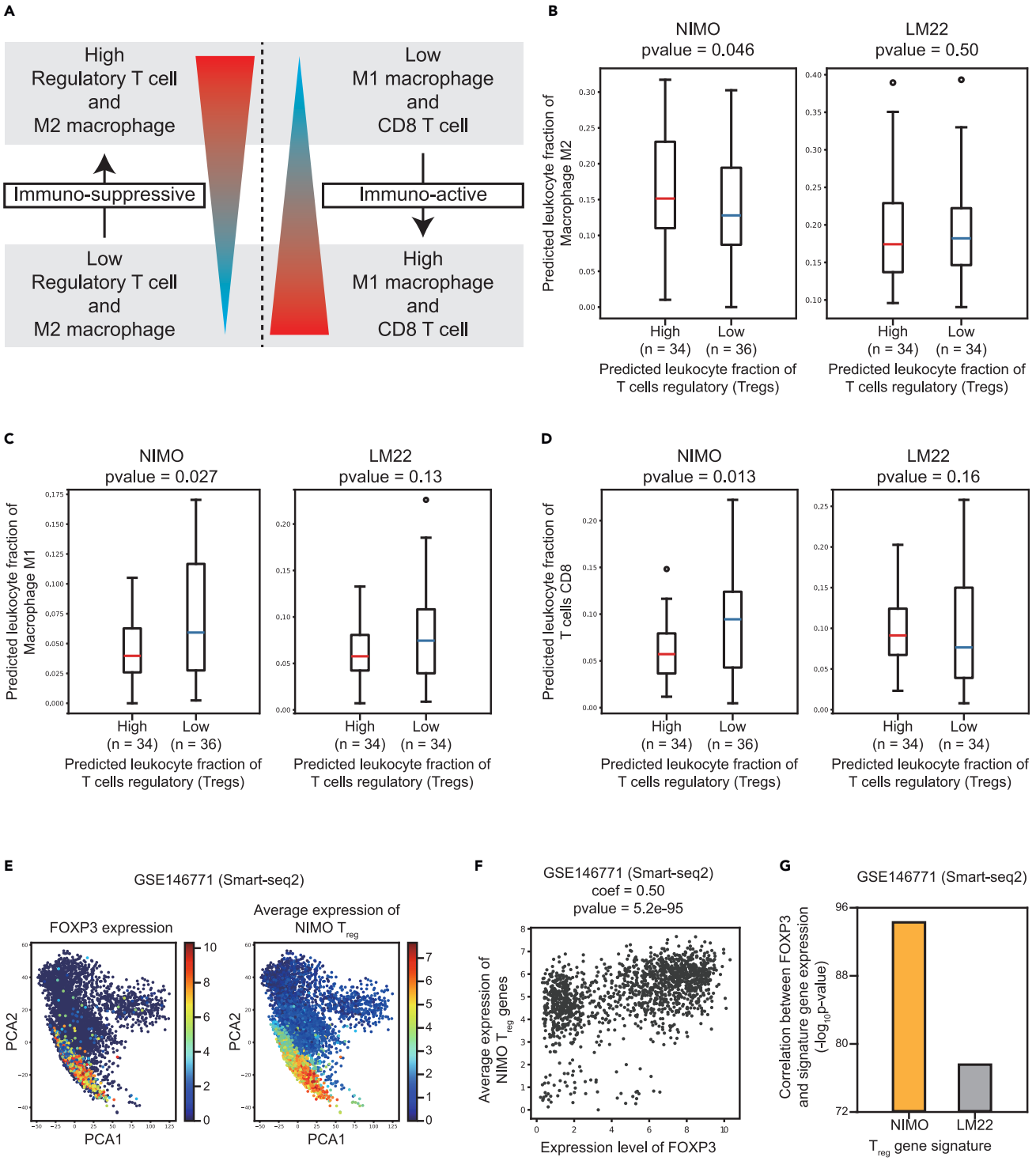


Figure 5. Regulatory T cell signatures inferred from the NIMO-based approach correspond with known biological characteristics

(A) Overview of crosstalk between regulatory T cells and M1 and M2 macrophages and CD8 T cells.

(B–D) Comparison of crosstalk between regulatory T cells and (B) M2 macrophages, (C) M1 macrophages, and (D) CD8 T cells using cell fractions inferred from the NIMO or LM22 signatures. Patients with CRC were split into high (upper 30%) and low (lower 30%) immune cell groups. Statistical significance was measured using the Mann-Whitney U test.

(E–G) Comparison of FOXP3 expression and the average expression of regulatory T cell genes from NIMO and LM22 signatures in an external single-cell RNA sequencing dataset.

(legend continued on next page)

single-cell RNA sequencing data. Our results suggest that with the use of appropriate immune cell signature genes, clinicians can better interpret the patient's immune status, which can be leveraged for predicting the clinical outcome of the patient.

Our findings provide further evidence for the effectiveness of network expansion analysis using multiomics data to identify robust biomarkers. Specifically, our study demonstrated that using NIMO-based marker genes for immune cell deconvolution improved the prediction of CRC recurrence (Figure 1), highlighting the importance of selecting appropriate marker genes and conducting network expansion analysis with multiomics data. In line with these results, incorporating biological knowledge into the biomarker discovery process is beneficial for identifying biomarkers for different immune cell types,[35,36] predicting anticancer drug responses in patients with cancer,[20] and identifying novel genes associated with oral disorders.[37,38] Recent studies by Strum et al. and Cobos et al. also underscored the impact of signature gene selection on immune cell deconvolution performance, further emphasizing the importance of careful marker gene selection.[14,15] In summary, our results highlight the crucial role that immune cell signature gene selection plays in providing an accurate understanding of the immune context; these results also highlight the potential utility of the NIMO approach for discovering such signature genes.

CRC is among the top three most common cancers in the USA; therefore, there is an urgent need to identify biomarkers that enable early prediction of clinical outcomes.[27,39] We found that the immune context of a patient can be used to predict recurrence in CRC (Figure 2). Additionally, we found that incorporating both the patient's immune context and TNM stage improved the prediction of recurrence in our CRC dataset. We also observed that the immune context and clinical features were variable even within groups of patients with and without recurrence (Figure 4), suggesting the existence of high heterogeneity within patients with CRC. Indeed, while TNM stage has been previously reported to be associated with cancer prognosis and metastasis in CRC,[27] patients displayed various outcomes even when they were assigned to the same TNM category.[3] These observations suggest that a combination of several different biomarkers is required for improving the survival of patients with CRC.[27] Because of the increasing evidence that the tumor microenvironment is closely associated with tumor development,[5] further clarification is required for the role of immune cells in cancer recurrence.

In this study, we chose CIBERSORT due to its ability to provide references for diverse immune cell types (22 types).[40] The quantification of immune cells in patients with cancer remains a crucial research topic for understanding cancer development in the context of the immune system. To this end, cell deconvolution methods have been developed to accurately quantify immune cell proportions in tumors, providing a cost-effective and scalable approach to analyze the immune context.[40] While various deconvolution methods have been developed[12,41,42] and their performances have been compared using benchmarking frameworks,[40] selecting the appropriate method for a given application requires careful consideration.

While we show that the NIMO-based approach can improve the prediction of recurrence in patients with CRC, it is still unclear whether our approach can be applied to predicting other clinical outcome-related factors, such as prognosis or immunotherapy response. Moreover, it is unknown whether NIMO can be used for analyses based on single-cell RNA sequencing. Reassuringly, we observed that the expression of T_{reg} cell signature genes found using the NIMO-based approach was in concordance with that of the T_{reg} cell-specific marker FOXP3 in single-cell RNA sequencing data of CRC samples (Figures 5E–5G). As more single-cell RNA sequencing data become available, we expect that the discovery of cell-type markers will be facilitated. An interesting research strategy would be to apply network-based approaches to single-cell RNA sequencing data to supplement the limitations of the technique in identifying cell-type markers.

The immune cell signatures identified by the NIMO approach are highly enriched in known immune cell markers (Figure S3) and help predict CRC recurrence (Figures 1 and 2). However, linking DNA methylation at enhancers to gene expression requires caution because DNA methylation at enhancers does not always guarantee regulation of the expression of their target genes. For example, Andersson et al.[43] reported that only half of the enhancers defined by cap analysis of gene expression (CAGE) sequencing (CAGE-seq) were tightly linked to their closest genes. In addition, our study observed a significant negative correlation between expression and methylation levels for 39% of the NIMO immune cell signatures (Figure S4). Currently, the lack of datasets with simultaneous expression and methylation sequencing for sorted immune cells presents a limitation. We anticipate that using a matched dataset with the NIMO approach could lead to more precise recurrence predictions in the future. To achieve an elaborate mapping of methylation with expression, it is necessary to infer enhancer-gene relationships using a dedicated physical interaction map between gene and regulatory regions that can be assayed by chromosome conformation capture (3C).[44,45]

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Sanguk Kim (sukim@postech.ac.kr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Methylation profiles of bulk tumors from patients with CRC from the SMC cohort have been deposited in The European Nucleotide Archive under the accession

(E) Principal-component analysis (PCA) plot using whole-transcriptome single-cell RNA sequencing data of patients with CRC (GEO: GSE146771). Colors depict the expression level of (left) FOXP3 or (right) the average expression of regulatory T cell markers from the NIMO signature.

(F) Correlation between FOXP3 expression and the average expression of regulatory T cell markers from the NIMO signatures. The correlation coefficient and significance were measured using Spearman correlation.

(G) Comparison of the significance of the correlation between FOXP3 expression and the average expression of regulatory T cell markers of NIMO and signature-based regulatory T cell markers.

code PRJEB50005 (https://www.ebi.ac.uk/ena/browser/view/PRJEB50005).[46] The raw data of bulk RNA sequencing from the SMC cohort have been deposited at the European Genome-Phenome Archive, which is hosted by the Geninus WTS Data Access Committee under accession code EGAD00001006985.[47] Immune cell compositions identified by the NIMO approach in patients with CRC have been provided in the GitHub repository (https://github.com/SBIlab/SGI_cancer_recurrence_NIMO/).[48] We used single-cell RNA sequencing data of colon cancer from Gene Expression Omnibus (GEO: GSE146771).[49]

Study approval

This study was approved by the institutional review boards of the SMC (approval no. SMC 2018-04-074-004). Written informed consent was obtained from all subjects. All experimental methods complied with the Helsinki Declaration.

Preprocessing of gene expression and methylation data

We aligned RNA sequencing reads to the human reference genome (GRCh38) using STAR aligner.[50] Read counts were used to obtain normalized gene expression levels using the RSEM package.[51] To quantify immune cell proportions in each patient, we used transcripts per million (TPM)-normalized RNA sequencing values.

GEO: GSE107422 [25] was curated as an independent cohort of patients with CRC with recurrence. We used processed TPM values as gene expression levels. When multiple gene IDs were annotated to a single gene symbol, we took the maximum TPM value of the gene IDs to map the gene ID to the gene symbol.

To preprocess the methylation data, we aligned the sequencing reads from bisulfite-treated samples and performed methylation calls using Bismark[52] with the default option. The alignment was performed on the human reference genome (hg19) using Illumina annotation. To quantify the immune cell proportions in each patient using methylation data, we used the β value, which is a fraction of the methylation per site. The β value was quantified using the R package bsseq.[53]

Identification of immune cell signature genes and inference of immune cell proportions in patients with CRC

To identify robust immune cell signature genes, we first curated transcriptome-based (T signature) and methylation-based (M signature) immune cell signatures from Newman et al. (LM22)[12] and Chakravarthy et al.,[13] respectively. We mapped the CpG sites in M signatures to their most proximal gene regulatory elements (i.e., promoters and enhancers), with promoter and enhancer regions downloaded from the FANTOM5 project.[43,54,55] The NIMO signature included both genes in the T signature and those mapped to the M signature. To identify signature genes with the NIMO approach, we used the STRING functional protein interaction network,[56] which was comprised of 13,824 nodes and 323,774 edges. The genes that were included in the signature met the following two criteria: (1) a first neighbor of T and M signature genes in the protein interaction network and (2) a gene included in the original LM22 signature genes. Source codes to identify immune cell signature genes are available at the following GitHub repository: https://github.com/SBIlab/SGI_cancer_recurrence_NIMO.[48]

To confirm whether the TM signature genes showed a correlation between expression and the DNA methylation levels of their proximal CpG sites, we utilized 359 TCGA COAD and READ samples.[57] The DNA methylation levels of the samples were measured by the Illumina Human Methylation 450 platform. The Spearman correlation coefficient was measured between gene expression levels and the most proximal CpG methylation levels of T and T-TM signatures. The p values of the correlation were adjusted by the Bonferroni method, and adjusted p values lower than 0.05 were considered significant. The proportions of negative significant, positive significant, and nonsignificant correlations of gene-CpG pairs were compared between the T signature and the T-TM signature. The significance of proportional bias was measured by using Fisher's exact test.

To infer the immune cell proportions of patients with CRC using different immune cell signatures (i.e., LM22, NIMO signatures), we used the transcriptome data from the patients and applied the CIBERSORT[12] deconvolution method using each immune cell signature. We used the reference LM22 immune cell expression data that were provided in the original publication. The deconvolution of CRC patient data using transcriptome- or NIMO-based signatures returned proportions of 22 immune cell types, including B cells (naive, memory, and plasma cells); monocytes; macrophages (M0, M1, and M2); T cells (CD4 memory resting, CD4 memory activated, CD4 naive, Treg, follicular helper, CD8, and $\gamma\delta$); natural killer (NK) cells (resting and activated); mast cells (resting and activated); dendritic cells (resting and activated); neutrophils; and eosinophils. For deconvolution using a methylation-based signature, we used methylCIBERSORT[13] to infer immune cell proportions using methylation data from patients with CRC in the SMC cohort.

Enrichment analysis of immune cell biomarkers for the TM signature

To validate the network expansion of the NIMO approach to improve the detection of markers of immune cell type, we used two cell biomarker databases, CellMarker[23] and PanglaoDB.[24] We selected 19 and 21 cell types in CellMarker and PanglaoDB, respectively, to match with LM22 cell types. First, we compared the number of immune cell markers covered by the NIMO signature genes with and without network expansion. The enrichment significance was measured using the hypergeometric test. The p values were corrected using the Bonferroni correction.

Recurrence prediction in the SMC cohort

We used cross-validation to evaluate the performance of immune cell proportions and/or clinical features in predicting recurrence. In detail, 80% of patients were divided into a training set, 20% of patients were divided into a test set, and different training and test sets were used for 500 iterations. We used an ensemble of ML models that included support vector machine, random forest, and extra tree models. The ensemble model was constructed by using a soft voting classifier that predicts recurrence by summing the predicted probability. All of the ML models were used with the sklearn[58] package in Python. To optimize each ML method, we used 5-fold cross-validation within the training set to select the best hyperparameters. The best hyperparameter for the support vector machine model was selected using the "rbf" kernel with kernel coefficient (γ). The values of regularization parameter (C) hyperparameters were tested at 0.001, 0.01, 1, 2, 3, 4, 5, 10, 20, 50, and 100. For the random forest model and extra tree classifier, the number of trees (n_estimators) was selected from 500, 100, and 2,000 trees. For logistic regression, we selected the regularization parameter C within 0.1 and 1 in 0.1 intervals. Final recurrence prediction for a patient was performed by voting from four different ML models using the VotingClassifier function from the sklearn Python package.

Recurrence prediction in the GEO: GSE107422 cohort

To validate NIMO markers in the independent GEO: GSE107422 cohort,[25] we performed cross-validation to evaluate the performance of immune cell proportions in predicting recurrence. Because of the unbalanced labels of the cohort, which was composed of 38 patients with recurrence and 72 patients without recurrence, balanced training and test sets were prepared. For recurrence patients, 80% of the patients were divided into a training set, 20% of the patients were divided into a test set, and the same number of patients without recurrence were randomly divided into training and test sets. Different training and test sets were used for 500 iterations.

Disease-free survival (DFS) analysis

To compare DFS between patients with predicted recurrence and nonrecurrence, we assigned predicted labels to patients with leave-one-out cross-validation (LOOCV). The Kaplan-Meier curve of patients' DFS was plotted using lifelines (v.0.19.5), a python module. The statistical significance of the survival difference was calculated by the log rank test.

Stage-specific prediction

We conducted stage-by-stage recurrence prediction for patients from the SMC cohort. Since the number of patients with a recurrence in TNM stages I, II, III, and IV were 2, 5, 19, and 21, respectively, we used patients in stages III and IV for the recurrence prediction. To evaluate the performance of recurrence prediction in each stage, we used 80% of all patients randomly selected as the ML training set, and among the remaining 20% of patients, only patients belonging to the corresponding stage were used as part of the test set.

Statistics

Statistical differences in predictive performances, immune cell proportions, and expression levels were measured by the Mann-Whitney U test. The correlation between the average expression of signature genes and a biomarker gene was measured by Spearman correlation. The significance of DEGs was corrected by multiple testing correction using the Benjamini-Hochberg procedure. Figure legends indicate the detailed methods of statistical analysis.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2023.100736.

ACKNOWLEDGMENTS

We thank all of the members of the Kim laboratory for helpful discussions. This work was partially supported by grants from the Korean National Research Foundation (2021R1A2B5B01001903, 2020R1A6A1A03047902, and 2020R1A2C1102210), the Ministry of Oceans and Fisheries ("Omics based on fishery disease control technology development and industrialization" [20150242]), the IITP (2019-0-01906, Artificial Intelligence Graduate School Program, POSTECH), the Ministry of Science and ICT (MSIT, 2018M3C9A6017315), and the Korean government (Ministry of Science and ICT, Ministry of Trade, Industry and Energy, Ministry of Health & Welfare, Ministry of Food and Drug Safety, KMDF_PR_20200901_ 0018).

AUTHOR CONTRIBUTIONS

J. Kong, J. Kim, D.K., K.L., S.K.H., I.K., Y.B.C., D.P., and S.K. conceptualized the research. J. Kong, J. Kim, D.K., K.L., J.L., S.K.H., I.K., S.L., M.P., and S.K. conducted the analysis. Funding was acquired by Y.B.C., D.P., and S.K. Patient curation was performed by S.S., W.Y.L., S.H.Y., H.C.K., H.K.H., Y.B.C., and D.P. Writing of the original draft was performed by J. Kong, J. Kim, D.K., K.L., S.K.H., I.K., Y.B.C., D.P., and S.K. J. Kong, J. Kim, D.K., K.L., J.L., and S.K.H. contributed equally to the analysis. The authorship order was determined by J.H.K. and S.K. initiating the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 27, 2022
Revised: November 21, 2022
Accepted: March 28, 2023
Published: April 20, 2023

REFERENCES

1. Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A., Smith, R.A., and Jemal, A. (2020). Colorectal cancer statistics, 2020. CA. Cancer J. Clin. *70*, 145–164. https://doi.org/10.3322/caac.21601.

2. Galon, J., Mlecnik, B., Bindea, G., Angell, H.K., Berger, A., Lagorce, C., Lugli, A., Zlobec, I., Hartmann, A., Bifulco, C., et al. (2014). Towards the introduction of the "Immunoscore" in the classification of malignant tumours. J. Pathol. *232*, 199–209. https://doi.org/10.1002/path.4287.

3. Nagtegaal, I.D., Quirke, P., and Schmoll, H.J. (2011). Has the new TNM classification for colorectal cancer improved care? Nat. Rev. Clin. Oncol. *9*, 119–123. https://doi.org/10.1038/nrclinonc.2011.157.

4. Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T.T., Berger, M.D., et al. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet *391*, 2128–2139. https://doi.org/10.1016/S0140-6736(18)30789-X.

5. Angell, H.K., Bruni, D., Barrett, J.C., Herbst, R., and Galon, J. (2020). The immunoscore: colon cancer and beyond a C. Clin. Cancer Res. *26*, 332–339. https://doi.org/10.1158/1078-0432.CCR-18-1851.

6. Mlecnik, B., Tosolini, M., Kirilovsky, A., Berger, A., Bindea, G., Meatchi, T., Bruneval, P., Trajanoski, Z., Fridman, W.H., Pagès, F., and Galon, J. (2011). Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. J. Clin. Oncol. *29*, 610–618. https://doi.org/10.1200/JCO.2010.30.5425.

7. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science *313*, 1960–1964. https://doi.org/10.1126/science.1129139.

8. Fridman, W.H., Zitvogel, L., Sautès-Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. Nat. Rev. Clin. Oncol. *14*, 717–734. https://doi.org/10.1038/nrclinonc.2017.101.

9. Fridman, W.H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer *12*, 298–306. https://doi.org/10.1038/nrc3245.

10. Mlecnik, B., Bindea, G., Pagès, F., and Galon, J. (2011). Tumor immunosurveillance in human cancers. Cancer Metastasis Rev. *30*, 5–12. https://doi.org/10.1007/s10555-011-9270-7.

11. Bindea, G., Mlecnik, B., Fridman, W.H., Pagès, F., and Galon, J. (2010). Natural immunity to cancer in humans. Curr. Opin. Immunol. *22*, 215–222. https://doi.org/10.1016/j.coi.2010.02.006.

12. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457. https://doi.org/10.1038/nmeth.3337.

13. Chakravarthy, A., Furness, A., Joshi, K., Ghorani, E., Ford, K., Ward, M.J., King, E.v., Lechner, M., Marafioti, T., Quezada, S.A., et al. (2018). Pan-cancer deconvolution of tumour composition using DNA methylation. Nat. Commun. *9*, 3220. https://doi.org/10.1038/s41467-018-05570-1.

14. Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics *35*, i436–i445. https://doi.org/10.1093/bioinformatics/btz363.

15. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E., Mestdagh, P., and de Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat. Commun. *11*, 5650. https://doi.org/10.1038/s41467-020-19015-1.

16. Nirmal, A.J., Regan, T., Shih, B.B., Hume, D.A., Sims, A.H., and Freeman, T.C. (2018). Immune cell gene signatures for profiling the microenvironment of solid tumors. Cancer Immunol. Res. *6*, 1388–1400. https://doi.org/10.1158/2326-6066.CIR-18-0342.

17. Jing, Y., Liu, J., Ye, Y., Pan, L., Deng, H., Wang, Y., Yang, Y., Diao, L., Lin, S.H., Mills, G.B., et al. (2020). Multi-omics prediction of immune-related adverse events during checkpoint immunotherapy. Nat. Commun. *11*, 4946. https://doi.org/10.1038/s41467-020-18742-9.

18. Webber, J.T., Kaushik, S., and Bandyopadhyay, S. (2018). Integration of tumor genomic data with cell lines using multi-dimensional network modules improves cancer pharmacogenomics. Cell Syst. *7*, 526–536.e6. https://doi.org/10.1016/j.cels.2018.10.001.

19. Sharifi-Noghabi, H., Zolotareva, O., Collins, C.C., and Ester, M. (2019). MOLI: multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics *35*, i501–i509. https://doi.org/10.1093/bioinformatics/btz318.

20. Kong, J., Lee, H., Kim, D., Han, S.K., Ha, D., Shin, K., and Kim, S. (2020). Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. Nat. Commun. *11*, 5485. https://doi.org/10.1038/s41467-020-19313-8.

21. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., and Barabási, A.L. (2015). Uncovering disease-disease relationships

through the incomplete interactome. Science *347*, 1257601. https://doi.org/10.1126/science.1257601.

22. Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. *12*, 56–68. https://doi.org/10.1038/nrg2918.
23. Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019). CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. *47*, D721–D728. https://doi.org/10.1093/nar/gky900.
24. Franzén, O., Gan, L.M., and Björkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database *2019*. https://doi.org/10.1093/database/baz046.
25. Kim, S.K., Kim, S.Y., Kim, C.W., Roh, S.A., Ha, Y.J., Lee, J.L., Heo, H., Cho, D.H., Lee, J.S., Kim, Y.S., and Kim, J.C. (2019). A prognostic index based on an eleven gene signature to predict systemic recurrences in colorectal cancer. Exp. Mol. Med. *51*, 1–12. https://doi.org/10.1038/s12276-019-0319-y.
26. Lee, Y.S., Song, S.J., Hong, H.K., Oh, B.Y., Lee, W.Y., and Cho, Y.B. (2020). The FBW7-MCL-1 axis is key in M1 and M2 macrophage-related colon cancer cell progression: validating the immunotherapeutic value of targeting PI3Kγ. Exp. Mol. Med. *52*, 815–831. https://doi.org/10.1038/s12276-020-0436-7.
27. Koncina, E., Haan, S., Rauh, S., and Letellier, E. (2020). Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. Cancers *12*, 319. https://doi.org/10.3390/cancers12020319.
28. Testa, U., Pelosi, E., and Castelli, G. (2018). Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. Med. Sci. *6*, 31. https://doi.org/10.3390/medsci6020031.
29. Phipps, A.I., Limburg, P.J., Baron, J.A., Burnett-Hartman, A.N., Weisenberger, D.J., Laird, P.W., Sinicrope, F.A., Rosty, C., Buchanan, D.D., Potter, J.D., and Newcomb, P.A. (2015). Association between molecular subtypes of colorectal cancer and patient survival. Gastroenterology *148*, 77–87.e2. https://doi.org/10.1053/j.gastro.2014.09.038.
30. Sinicrope, F.A., Shi, Q., Smyrk, T.C., Thibodeau, S.N., Dienstmann, R., Guinney, J., Bot, B.M., Tejpar, S., Delorenzi, M., Goldberg, R.M., et al. (2015). Molecular markers identify subtypes of stage III colon cancer associated with patient outcomes. Gastroenterology *148*, 88–99. https://doi.org/10.1053/j.gastro.2014.09.041.
31. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356. https://doi.org/10.1038/nm.3967.
32. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. *24*, 1550–1558. https://doi.org/10.1038/s41591-018-0136-1.
33. Beyer, M., Mallmann, M.R., Xue, J., Staratschek-Jox, A., Vorholt, D., Krebs, W., Sommer, D., Sander, J., Mertens, C., Nino-Castro, A., et al. (2012). High-Resolution transcriptome of human macrophages. PLoS One *7*, e45466. https://doi.org/10.1371/journal.pone.0045466.
34. Najafi, M., Farhood, B., and Mortezaee, K. (2019). Contribution of regulatory T cells to cancer: a review. J. Cell. Physiol. *234*, 7983–7993. https://doi.org/10.1002/jcp.27553.
35. Lin, J., Yu, M., Xu, X., Wang, Y., Xing, H., An, J., Yang, J., Tang, C., Sun, D., and Zhu, Y. (2020). Identification of biomarkers related to CD8+ T cell infiltration with gene co-expression network in clear cell renal cell carcinoma. Aging *12*, 3694–3712. https://doi.org/10.18632/aging.102841.
36. Klimm, F., Toledo, E.M., Monfeuga, T., Zhang, F., Deane, C.M., and Reinert, G. (2020). Functional module detection through integration of single-cell RNA sequencing data with protein–protein interaction networks. BMC Genom. *21*, 756. https://doi.org/10.1186/s12864-020-07144-2.
37. Han, S.K., Kong, J., Kim, S., Lee, J.H., and Han, D.H. (2019). Exomic and transcriptomic alterations of hereditary gingival fibromatosis. Oral Dis. *25*, 1374–1383. https://doi.org/10.1111/odi.13093.
38. Choi, H., Lee, K., Kim, D., Kim, S., and Lee, J.H. (2022). The implication of holocytochrome c synthase mutation in Korean familial hypoplastic amelogenesis imperfecta. Clin. Oral Investig. *26*, 4487–4498. https://doi.org/10.1007/s00784-022-04413-0.
39. Tanoue, L.T. (2012). Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. Year Bk. Pulm. Dis. *2012*, 62–64. https://doi.org/10.1016/j.ypdi.2012.01.043.
40. Jin, H., and Liu, Z. (2021). A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. Genome Biol. *22*, 102. https://doi.org/10.1186/s13059-021-02290-6.
41. Liebner, D.A., Huang, K., and Parvin, J.D. (2014). MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. Bioinformatics *30*, 682–689. https://doi.org/10.1093/bioinformatics/btt566.
42. Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J.C., Rodig, S., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. *17*, 174. https://doi.org/10.1186/s13059-016-1028-7.
43. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461. https://doi.org/10.1038/nature12787.
44. Dekker, J. (2006). The three "C" s of chromosome conformation capture: controls, controls, controls. Nat. Methods *3*, 17–21. https://doi.org/10.1038/nmeth823.
45. Downes, D.J., Beagrie, R.A., Gosden, M.E., Telenius, J., Carpenter, S.J., Nussbaum, L., de Ornellas, S., Sergeant, M., Eijsbouts, C.Q., Schwessinger, R., et al. (2021). High-resolution targeted 3C interrogation of cis-regulatory element organization at genome-wide scale. Nat. Commun. *12*, 531. https://doi.org/10.1038/s41467-020-20809-6.
46. Kim, D., Kim, J., Lee, J., Han, S.K., Lee, K., Kong, J., Kim, Y.J., Lee, W.Y., Yun, S.H., Kim, H.C., et al. (2022). Deconvolution of bulk tumors into distinct immune cell states predicts colorectal cancer recurrence. iScience *25*, 105392. https://doi.org/10.1016/j.isci.2022.105392.
47. Park, D.. RNA-seq data from Korean CRC samples – European Genome-phenome Archive (EGA). https://ega-archive.org/datasets/EGAD00001006985.
48. SBIlab (2023). SBIlab/SGI_cancer_recurrence_NIMO: v1.0.0. https://doi.org/10.5281/ZENODO.7652592.
49. Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. Cell *181*, 442–459.e29. https://doi.org/10.1016/j.cell.2020.03.048.
50. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
51. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.
52. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
53. Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. *13*, R83. https://doi.org/10.1186/gb-2012-13-10-R83.
54. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., and Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable

modular perturbations across complex diseases. Nat. Methods *13*, 366–370. https://doi.org/10.1038/nmeth.3799.

55. FANTOM Consortium and the RIKEN PMI and CLST DGT, Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470. https://doi.org/10.1038/nature13182.

56. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613. https://doi.org/10.1093/nar/gky1131.

57. Cancer Genome Atlas Network, Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337. https://doi.org/10.1038/nature11252.

58. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*.