



# Computational Tumor Infiltration Phenotypes Enable the Spatial and Genomic Analysis of Immune Infiltration in Colorectal Cancer

Henrik Failmezger<sup>1</sup>, Natalie Zwing<sup>2</sup>, Achim Tresch<sup>3,4,5\*</sup>, Konstanty Korski<sup>2\*</sup> and Fabian Schmich<sup>1\*</sup>

<sup>1</sup> Data Science, Pharma Research and Early Development, Roche Innovation Center Munich, Penzberg, Germany, <sup>2</sup> Early Biomarker Development Oncology, Pharma Research and Early Development, Roche Innovation Center Munich, Penzberg, Germany, <sup>3</sup> Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany, <sup>4</sup> Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany, <sup>5</sup> Center for Data and Simulation Science, University of Cologne, Cologne, Germany

## OPEN ACCESS

### Edited by:

Samata Kakkad,  
Merck, United States

### Reviewed by:

Rolf Turner,  
The University of Auckland,  
New Zealand  
Shikhar Uttam,  
University of Pittsburgh, United States

### \*Correspondence:

Achim Tresch  
achim.tresch@uni-koeln.de  
orcid.org/0000-0002-4146-637  
Konstanty Korski  
konstanty.korski@roche.com  
Fabian Schmich  
fabian.schmich@roche.com  
orcid.org/0000-0003-1604-440

### Specialty section:

This article was submitted to  
Cancer Imaging and Image-directed  
Interventions,  
a section of the journal  
Frontiers in Oncology

**Received:** 15 April 2020

**Accepted:** 10 February 2021

**Published:** 15 March 2021

### Citation:

Failmezger H, Zwing N, Tresch A,  
Korski K and Schmich F (2021)  
Computational Tumor Infiltration  
Phenotypes Enable the Spatial and  
Genomic Analysis of Immune  
Infiltration in Colorectal Cancer.  
*Front. Oncol.* 11:552331.  
doi: 10.3389/fonc.2021.552331

Cancer immunotherapy has led to significant therapeutic progress in the treatment of metastatic and formerly untreatable tumors. However, drug response rates are variable and often only a subgroup of patients will show durable response to a treatment. Biomarkers that help to select those patients that will benefit the most from immunotherapy are thus of crucial importance. Here, we aim to identify such biomarkers by investigating the tumor microenvironment, i.e., the interplay between different cell types like immune cells, stromal cells and malignant cells within the tumor and developed a computational method that determines spatial tumor infiltration phenotypes. Our method is based on spatial point pattern analysis of immunohistochemically stained colorectal cancer tumor tissue and accounts for the intra-tumor heterogeneity of immune infiltration. We show that, compared to base-line models, tumor infiltration phenotypes provide significant additional support for the prediction of established biomarkers in a colorectal cancer patient cohort ( $n = 80$ ). Integration of tumor infiltration phenotypes with genetic and genomic data from the same patients furthermore revealed significant associations between spatial infiltration patterns and common mutations in colorectal cancer and gene expression signatures. Based on these associations, we computed novel gene signatures that allow one to predict spatial tumor infiltration patterns from gene expression data only and validated this approach in a separate dataset from the Cancer Genome Atlas.

**Keywords:** immunotherapy, machine learning, colorectal cancer, spatial statistics, immune response

## INTRODUCTION

Cancer immunotherapy is the most promising therapy for many metastatic and formerly untreatable tumors. However, often only a subgroup of patients will benefit from its application. Biomarkers are important predictors of patients' response to a treatment and, moreover, offer new insights into drug mechanisms of action (1, 2). There are different types of biomarkers: For instance, over-expression of genes like PDL1, mutations in specific genes like KRAS or BRAF or impaired DNA mismatch repair (microsatellite instability, MSI) have all been shown to be

informative biomarkers in the context of immunotherapy for colorectal cancer (CRC) patients (2). While recently, tumor mutation burden (TMB), i.e., the number of mutations per megabase has been established as another genetic biomarker (3, 4), the field is currently shifting its focus toward the tumor microenvironment (TME) (5–8). The TME is characterized by the complex interaction between malignant tumor cells, immune cells (e.g., CD8+ or CD4+ T cells), as well as stromal cells. It has been shown that the presence of immune cells in malignant tumors is predictive of the success of certain immunotherapies (5, 9). Abundances of immune cells are either determined by gene expression signatures from bulk RNA sequencing (10), or based on immunohistochemical (IHC) staining of tissue samples (11) and subsequent calculation of the immune cell density, i.e., the number of immune cells divided by the tissue area. Traditionally, immune cell densities have been hand-annotated by a pathologist, whereas today, many such workflows have been automated and digitalized using methodology from the fields of Artificial Intelligence and Computer Vision (1, 6, 12).

However, neither gene expression based immune cell abundances, nor immune cell densities derived from IHC provide insight into whether T cells are effectively infiltrating a tumor or whether they are blocked outside of the tumor and concentrated within the stromal tissue. Recently, three major patterns of tumor immune infiltration have been described: *deserted* (no to low T cell abundance), *excluded* (T cells and tumor cells occupy disjoint spatial areas) and *inflamed* (T cells and tumor cells spatially colocalize) (13). It is hypothesized that the spatial arrangement of T cells and tumor cells and in particular the proximity of cytotoxic CD8+ T cells to malignant tumor cells is strongly affecting the immune response (14, 15).

While *deserted* tumors can be fully characterized based on T cell abundance, the distinction between *excluded* and *inflamed* tumors requires spatial pattern analysis. In this work, we present a novel computational method based on Ripley's L function (16) that, for the first time, is capable to successfully distinguish *excluded* and *inflamed* tumor infiltration phenotypes (TIPs) in CRC tissue (**Figure 1B**). Ripley's L function has previously been used to characterize the patterning of stromal cells in breast cancer (6) and the patterning of CD8+ T cells and tumor cells in pancreatic cancer (14).

We observed that the patterning of T cells in many tumors showed substantial local heterogeneity across the tissue sample. In order to account for this spatial variation of infiltration phenotypes, we tessellated the sample area into hexagonal tiles (edge length = 375  $\mu\text{m}$ ) and calculated the co-localization of T cells and tumor cells per tile. Each tile's specific shape of the L function then served as the input features to a logistic fused lasso regression model (17, 18), that was subsequently trained to classify each tile to either infiltration phenotype (*inflamed* vs. *excluded*). In order to aggregate tile-level information to a quantitative biomarker on the sample level, we computed the respective proportion of tiles for each TIP across the tissue sample. In total, we characterized TIPs for 80 samples from CRC patients. We show that TIPs add valuable information to the prediction of established biomarkers like TMB and MSI and immune cell subpopulations, as defined by gene

expression signatures. Further, we developed a novel gene expression deconvolution scheme, accounting for differently sized populations of immune- and tumor cells within a sample, which allowed us to infer distinct gene expression signatures that predict the spatial tumor infiltration phenotypes of a sample. We validated the prediction of these phenotypes from gene signatures in the Cancer Genome Atlas.

## MATERIALS AND METHODS

### Histological Data

Surgical specimens from primary colorectal tumors were procured. They were collected from consented patients (informed consent) and under approval from the respective Institutional Review Board, National Ethics Committee, or equivalent agency. The samples had been fixed in formalin, embedded in paraffin and archived prior to shipment.

A total amount of 80 tissue specimens was obtained. The cells in the sections were stained by fluorescent immunohistochemistry with the markers Ki67+ (30-9, Ventana Medical Systems), CD3+/CD4+ (2GV6, Ventana Medical Systems) and CD3+/CD8+ (SP239, Spring Biosciences) (19). The slides were digitized with a Zeiss Axio Scan.Z1 at 20x magnification, resulting in a pixel resolution of 325 nm.

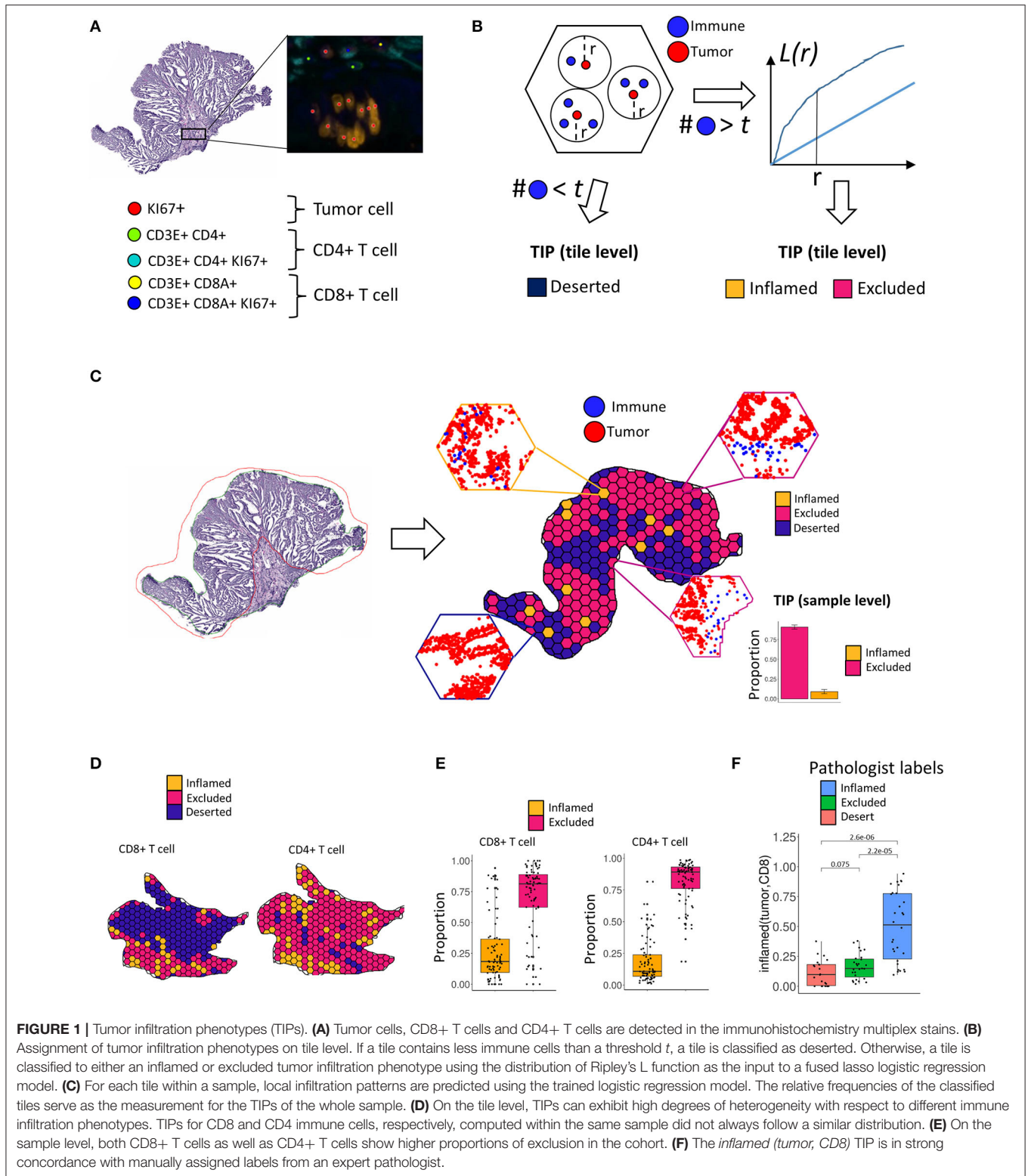
In order to determine the MSI status, the slides of the tissue samples were stained and assessed for the presence or absence of four mismatch repair (MMR) proteins, MLH1 (M1, Ventana), MSH2 (G219-1129, Ventana), MSH6 (SP93, Ventana), and PMS2 (A16-4, Ventana). Tumors with loss of one or more of the MMR proteins are considered MSI, whereas intact MMR staining is classified as MSS (20). 24% of the patients were MSI (**Table 1**) and most patients were at a late tumor stage (stage 3: 20%, stage 4: 64%) (**Table 1**). Tumor mutational burden was calculated as the number of mutations per megabase.

### Image Processing and Analysis

Tumor and immune cells were detected and classified by a proprietary machine learning algorithm based on color, intensity, texture, object shape. For each cell, we obtained its position and its cell type (**Figure 1A**): Tumor cells (stained by Ki67+), CD4 cells (stained by CD3+CD4+ or CD3+CD4+Ki67+), CD8 cells (stained by CD3+CD8A+ or CD3+CD8A+Ki67+). An expert pathologist, who also verified the results of the algorithm, manually annotated tumor regions. Only cells within the manually labeled tumor region were used for analysis. Furthermore, we excluded areas within each tile (hexagons, side length = 50  $\mu\text{m}$ ) with less than two cells in order to avoid artifacts caused by these regions, e.g., holes could bias the computation of the Ripley's L function (**Supplementary Figure 1A**).

### Genomic Data

The mutation dataset consisted of 373 cancer and immune related genes based on the gene panel of FoundationOne (21). Mutation data was available for 75 patients. For the mutation analysis, we kept only genes that had a mutation in at least 20% of the patients, resulting in a dataset of 19 genes (**Supplementary Table 6**).



For generating the RNAseq data for the tissue samples (22), genomic DNA and total RNA were purified from 10  $\mu$ m thick FFPET curls using the AllPrep DNA/RNA FFPE Kit (Qiagen

Cat No./ID: 80234). The Qubit instrument was used to assess the RNA samples for quality and quantity and the Agilent Bioanalyzer was used to determine the degradation of the RNA

**TABLE 1** | Characteristics of the dataset.

Characteristic	N = 80
<b>Age<sup>a</sup></b>	69 (60,77)
<b>Tumor grade</b>	
G1	9 (11%)
G2	45 (56%)
G3	26 (32%)
<b>Gender</b>	
Female	40 (50%)
Male	40 (50%)
<b>MMR status</b>	
MSI	19 (24%)
MSS	61 (76%)
<b>Tumor stage</b>	
I	2 (2.5%)
II	11 (14%)
III	16 (20%)
IV	51 (64%)

<sup>a</sup>Statistics presented: median (IQR).

samples (DV200 value). To further generate the sequencing library, the hybridization-based Illumina TruSeq RNA Access method was performed, with first preparation of the total RNA library and second library enrichment for coding RNA. Finally, normalized libraries were sequenced using the Illumina sequencing-by-synthesis platform, with a sequencing protocol of 50 bp paired-end sequencing and total read depth of 25M reads per sample (22). The data has been deposited as part of a previous publication in the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152395>). For further analyses, we selected from the gene expression data all genes that had no missing value in any patient resulting in RNAseq count data for 14,634 genes and 66 patients. Signatures for immune subpopulations were derived from Angelova et al. (23), while the effect size of association of a signature and a patient sample was calculated by gene set enrichment analysis using the gCMAP package (24, 25).

Gene expression data for the TCGA-COAD and TCGA-READ datasets from the Cancer Genome Atlas (TCGA) were retrieved by TCGAbiolinks (26). Tumors with the classification: *Splenic flexure of colon, Sigmoid colon, Descending colon, Rectosigmoid junction* were considered as left side tumors. Tumors with the classification: *Ascending colon, Transverse colon, Cecum, Hepatic flexure of colon* were considered as right side tumors.

### Pathway Mutation Score

Pathways were retrieved from the KEGG database (27). We selected pathways for which the overlap between the pathway genes and our mutation gene set (373 genes) was at least 40% which resulted in 14 pathways (Figure 4B).

The pathway mutation score  $m_{i,p}$  for a patient  $i$  and pathway  $p$  was calculated as:

$$m_{i,p} = \frac{\text{\#mutated genes for patient } i \text{ in pathway } p}{\text{\#genes in pathway } p}$$

### Spatial Statistics Measuring Interaction of Cells

Tumor tissue was tessellated into hexagonal tiles with a side length of 375  $\mu\text{m}$ . Based on spatial  $x$  and  $y$  coordinates of each cell, as derived from image processing of the imaged tissue, Ripley's L function was calculated for different combinations of cell types  $i$  and  $j$  within each tile.  $L_{ij}(r)$  is defined

as:  $L_{ij}(r) = \sqrt{\frac{K_{ij}(r)}{\pi}}$ , where

$K_{ij}(r) = \frac{1}{\lambda_j} * E \left[ t(u, r, X^{(j)}) | u \in X^{(i)} \right]$  (28),  $\lambda_j$  is the density of cell type  $j$  and

$t(u, r, X^{(j)}) = \sum_{k=1}^{n(x^{(j)})} 1\{0 < \|u - x_k\| \leq r\}$  is the number of cells of type  $j$  that lie within a distance  $r$  of the location  $u$  (28). In what follows,  $i$  always denotes tumor cells, and  $j$  either denotes CD8 or CD4 cells, i.e., we quantify the proximity of immune cells to tumor cells from the perspective of each tumor cell. The expectation value in Ripley's L function is estimated by empirical cell frequencies. Since the area under scrutiny however is finite, it is necessary to apply isotropic edge correction (28) to obtain an unbiased estimate.

### Calculation of Tumor Infiltration Phenotypes

In order to cast Ripley's L curves into the phenotypes *inflamed* vs. *excluded*, we trained a 1d logistic fused lasso regression model based on each tile's feature vector. Let  $S$  be the number of samples (tiles). Let  $y_s \in \{inflamed, excluded\}$ ,  $s = 1, \dots, S$  be the ground truth labels of the tiles. A feature vector  $x_{sr} = L_{sij}(r) - r$ , with  $r$  ranging over an equidistant grid of  $n$  values in the range of 1 to 338  $\mu\text{m}$ , is calculated for every tile. Here  $L_{sij}(r)$  is the observed L function in the tile for cell types  $i$  and  $j$  and  $r$  is the expected value of the L function under the assumption that there is independence between the two point patterns. The logistic fused lasso regression model is defined as:

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^n} -\frac{1}{S} \sum_{s=1}^S \left\{ y_s \left( \beta_0 + \beta^T x_s \right) - \log \left( 1 + e^{\beta_0 + \beta^T x_s} \right) \right\} + \lambda_1 \sum_{r=1}^{n-1} |\beta_{r+1} - \beta_r| + \lambda_2 \sum_{r=1}^n |\beta_r| \quad (17, 18)$$

The ground truth labels  $y_s$ , required to fit the logistic regression model, were annotated by an expert pathologist. The penalty term  $\lambda_1$  in the model enforces smoothness of neighboring coefficients, i.e., points on the Ripley's L curve, and makes it particularly suitable for a situation, where we expect high degrees of correlation between consecutive points in this 1d-sequential feature space. The fused lasso regression model was fit on a balanced training set of  $n = 118$  tiles. As fluorescent immunohistochemistry multiplex stainings are difficult to analyze visually this ground truth data was created based on a KI67/CD8 duplex staining. We chose a side length of 375  $\mu\text{m}$  for the tiles, which ensured a sufficient number of tiles per tissue, but was also large enough for the pathologist to evaluate the patterning of the cells in the tile. A minimum of more than five immune cells was required for a pathologist to

reliably distinguish the infiltration patterns *inflamed* vs. *excluded*. Therefore, if the number of immune cells within a tile was smaller than or equal to five, we set its infiltration pattern to *deserted*.

Estimates of Ripley’s L function become unstable at radii close to the maximum tile size (375 μm). We therefore restricted our feature vector to values of r from 1 to 338 μm, the latter being about 90% of the tile size. The parameters λ<sub>1</sub> and λ<sub>2</sub> that regularize both the difference of consecutive coefficients as well the coefficients by itself were determined by five-fold cross validation (**Supplementary Figure 1B**). We use the notation *inflamed* (i,j) and *excluded* (i,j) to describe the TIPs, both on the tile level and on the whole tissue level. On the tile level the TIP *inflamed* (tumor, CD8), for instance, represents the classification result of the fused lasso regression model, whereas on the whole tissue level it represents the fraction of tiles classified *inflamed*. In order to make the calculations robust against the exact position of the tiles, we shifted the tile grid by 10% in all directions that are integer multiples of 90° and repeated the calculation of TIPs. The resulting four proportions and the original one are averaged to obtain the final TIP.

### Multivariate Regression Analysis for Biomarker Validation

We evaluated the predictive capacity of TIPs for established biomarkers and the gene expression signatures against a base model consisting of the covariates *Age, Tumor Stage > IV, CD8+ T cell density and CD4+ T cell density*. We selected *inflamed* (tumor, CD8) and *inflamed* (tumor, CD4) for evaluation.

In the process of building the extended model containing the TIPs, we first applied a univariate regression against an established CRC biomarker. We then checked if any of the tumor infiltration phenotypes had a Pearson correlation coefficient larger than 0.6 with another TIP. In this case, in order to avoid correlated co-variates, we selected the TIPs that performed better in the univariate regression. We added all uncorrelated TIPs to the base model, reduced the model by stepwise regression analysis (forward-backward selection) and tested if the advanced model significantly improved over the base model (likelihood ratio (LR) test).

### Gene Expression Deconvolution

Bulk RNAseq gene expression data is derived from a mixture of different cell types, most prominently immune cells and cancer cells in our application. Therefore, we implemented a gene expression deconvolution scheme in order to derive gene expression values with reduced bias from the cellular composition of the bulked measurement, i.e., highly imbalanced abundances of different cell types.

Let G be a set of genes and S a set of samples, let

$$X = (x_{g,s}) \in \mathbb{R}^{G \times S}$$

be the expression matrix normalized first by samples and afterwards by genes. The technical noise of gene expression data is following a poisson distribution (29) in which the variance equals the mean, thus we divided every gene by the square root of its mean (i.e., variance). By variance stabilizing each gene, we ensured that the linear approximation of all components is not

dominated by a few highly expressed genes and that the following convex optimization has a physical interpretation as a mixture of cells.

Let λ = (λ<sub>s</sub>; s ∈ S), 0 ≤ λ ≤ 1 be the vector of relative abundances of immune cells.

The task is to infer a prototype cancer cell expression profile C = (c<sub>g</sub>; g ∈ G) and an immune expression profile I = (i<sub>g</sub>; g ∈ G) such that X is approximated by a convex combination X̃ of I and C:

$$X \approx \tilde{X} = I * \lambda^T + C * (1 - \lambda)^T$$

This can be solved using constrained convex optimization:

$$\operatorname{argmin}_{I,C} (X - \tilde{X}(I, C))^2$$

Subject to the constraints:

$$i_g \geq 0; g \in G, \sum_g i_g = 1$$

$$c_g \geq 0; g \in G, \sum_g c_g = 1$$

### Gene Signatures

For every sample, s ∈ S the estimated value x̃<sub>g,s</sub>, s ∈ S reveals the closest approximation of the value x<sub>g,s</sub>, s ∈ S using only a mixture of the global cancer profile C and the global immune profile I. Genes that influence the spatial localization of immune cells (= the Tumor Infiltration Phenotype) should be independent from the x̃<sub>g,s</sub>, s ∈ S pattern. A measurement from a TIP a = (a<sub>s</sub>; s ∈ S) is thus associated to r<sub>g,s</sub> = x<sub>g,s</sub> - x̃<sub>g,s</sub>, r ∈ R. This association of a TIP a to the expression of a certain gene g can be estimated by linear regression:

$$a_s = \beta_g * r_{g,s}, r \in R$$

Genes that showed a significant association with a TIP (Bonferroni multiple testing correction) became part of this TIP’s gene expression signature.

Using the expression values of the genes in the signature as covariates, tumor infiltration phenotypes can be explained by gene expression alone. Therefore, we trained a support vector machine for the prediction of tumor infiltration phenotypes from gene signatures. The gamma and the cost parameter were tuned by a grid search (cost=4, gamma=0.0009765625).

### Other Statistical Methods and Data Availability

All statistical analysis was carried out using R ≤3.5.1 (R Core Team 2013). The R package spatstat was used for spatial analysis (28, 30, 31). The R package penalized was used for the fused lasso model (32). GGplot2 was used for visualization (33). The lmerTest package was used for the likelihood ratio test (34). We used the package lol for the stability selection approach (35).

CVXR was used for solving the gene expression deconvolution (36). TopGO was used for differential gene expression analysis

(37). The R-package `e1071` was used for the training of the support vector machine (38). Example sourcecode as well as all spatial point patterns can be found in supplementary data (**Supplementary File 2–6**).

## RESULTS

### Computation of Tumor Infiltration Patterns From IHC Stained Tissue

In order to determine local immune infiltration patterns in the tumor, CD8+ T cells, CD4+ T cells and tumor cells were first detected by an automated image processing pipeline (Methods, **Figure 1A**). Subsequently, in order to account for the spatial heterogeneity of immune infiltration, we separated the tumor sample into hexagonal tiles (Methods). For each tile, we computed Ripley's L function across pairs of tumor and immune cells within the tile (Methods). Ripley's L function is a cumulative function that determines, from the perspective of one type of cell, the number of neighbors of another type of cell within a certain distance  $r$  (**Figure 1B**, Methods). The shape of the L function is indicative of a segregating or an aggregating relationship between the two cell types.

After computing the L function, we classified each tile within the sample into local infiltration patterns (**Figure 1B**, Methods) using a fused lasso logistic regression model (Methods). The fused lasso regression model was trained on tiles that were manually labeled by a pathologist and resulted in a cross-validation accuracy of 85% (**Supplementary Figure 1B**, Methods).

The TIPs *inflamed (tumor, CD8)* and *inflamed (tumor, CD4)* represent patterns where tumor cells and immune cells are co-clustered, whereas the TIPs *excluded (tumor, CD8)* and *excluded (tumor, CD4)* represent patterns where immune cells are segregated from tumor cells.

We found that the tile-level distribution of TIPs varied substantially, indicating high degrees of intra sample heterogeneity. For the same sample, we also identified different spatial infiltration patterns for CD8+ T cells and CD4+ T cells, respectively (**Figure 1D**). On the sample level we detected *excluded (tumor, CD8)* and *excluded (tumor, CD4)* to be the most frequent TIPs in the cohort (**Figure 1E**). We compared the TIP *inflamed (tumor, CD8)* to immune infiltration phenotypes manually assigned by an expert pathologist based on inspection of the whole tissue sample. Samples labeled *inflamed* by the pathologist also showed significantly higher fractions ( $p = 2.2e-05$ ) of the TIP *inflamed (tumor, CD8)*, indicating strong concordance between predicted infiltration patterns and experts' ground truth annotations (**Figure 1F**).

### Tumor Infiltration Phenotypes Predict Established, Prognostic Biomarkers in Colorectal Cancer

Patient survival and response in colorectal cancer is well-known to be linked with the presence of certain prognostic biomarkers, like the microsatellite instability (MSI) status, tumor mutational burden (TMB) and the expression of immune cell signatures. We

investigated the association of TIPs with these biomarkers and evaluated if TIPs could act as their independent predictors.

We found a positive univariate association of the TIP *inflamed (tumor, CD8)* with the patients' microsatellite instability (MSI) status ( $p = 5.8e-07$ , **Figure 2A**). In order to quantify the observed association, we investigated, if TIPs provide orthogonal, non-redundant information for the prediction of MSI and tumor mutational burden (TMB) when compared to a baseline model consisting of the biomarkers CD8+ T cell density, CD4+ T cell density, Age and Tumor stage ( $\geq$  Stage 4). All computationally derived TIPs were added as covariates to the base model. Subsequently, the number of covariates was iteratively reduced using stepwise regression (Methods). For the prediction of MSI, *inflamed (tumor, CD8)* was kept as a covariate in the extended model (**Figure 2B**, **Supplementary Table 1**), improving the overall model significantly ( $p = 0.009$ ).

Similarly, the TIP *inflamed (tumor, CD8)* was associated positively with high TMB [ $p = 3.1e-06$ , a TMB threshold of 20 was used for the assignment to the TMB high or low group (39), **Figure 2C**]. When compared to the base model, the prediction of TMB was again significantly improved by adding *inflamed (tumor, CD8)* ( $p = 0.004$ , **Figure 2D**, **Supplementary Table 2**).

Further, we investigated the association of TIPs with subpopulations of tumor infiltrating lymphocytes. We calculated the effect size of three immune signatures [cytotoxic T cell, T helper cell 1 (TH1) and T helper cell 17 (TH17) (23)] from matched gene expression data using gene set enrichment analysis (see Methods). There was an association between *inflamed (tumor, CD8)* and the sign of cytotoxic and TH1 immune signature's effect size (Cytotoxic cells:  $p = 0.00012$ , **Figure 3A**, TH1:  $p = 1.4e-06$ , **Figure 3C**) and between *inflamed (tumor, CD4)* and the effect size of the TH17 signature ( $p = 0.007$ , **Figure 3E**).

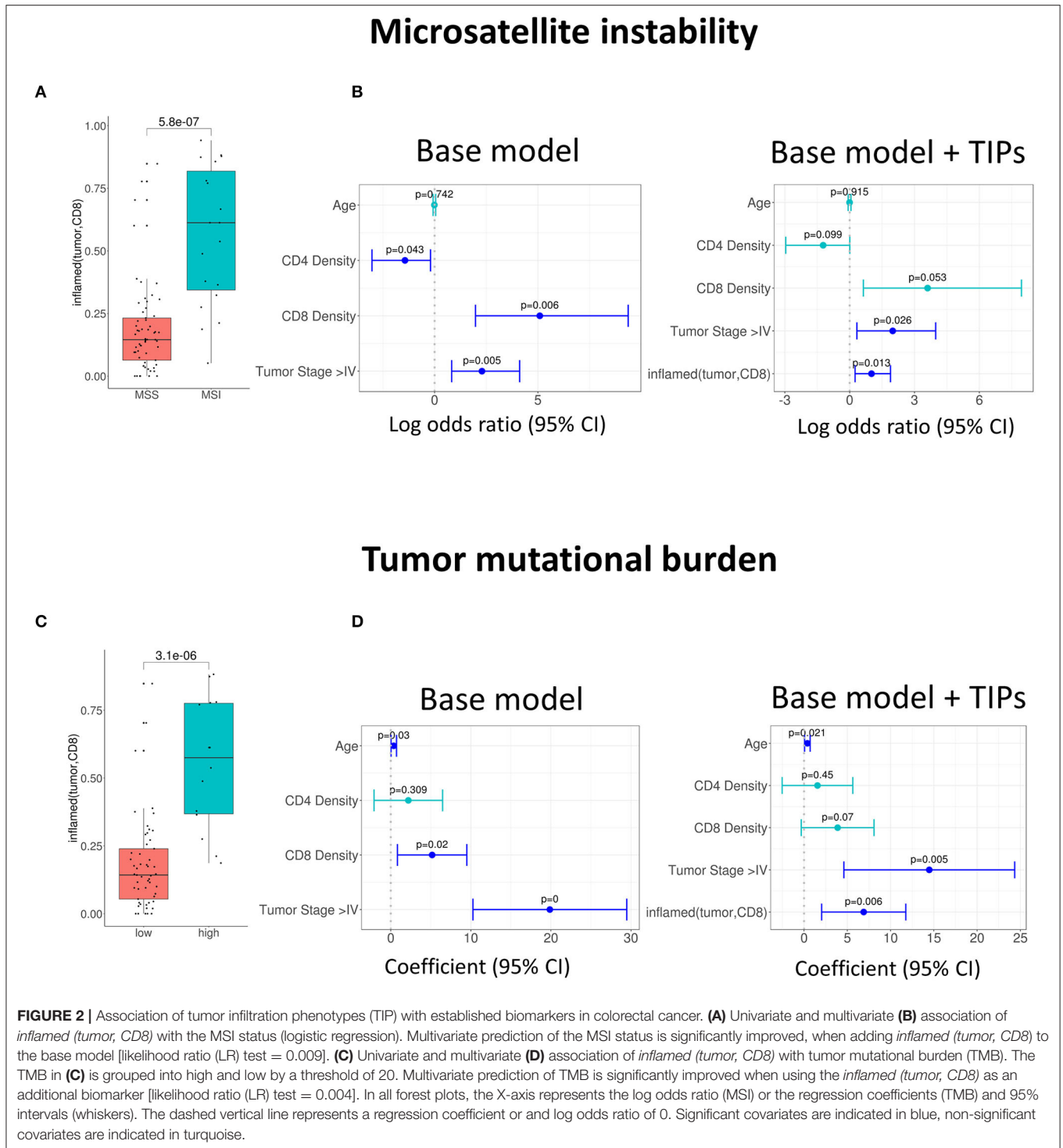
In order to quantify if the TIPs provide additional, non-redundant information for the prediction of the effect sizes of the three immune signatures, we again added them to the base model and reduced the number of covariates by stepwise regression (Methods).

For the prediction of the immune signature of cytotoxic cells, *inflamed (tumor, CD8)* was kept in the model. The addition of the TIP improved the model significantly ( $p = 5.01e-07$ , **Figure 3B**, **Supplementary Table 3**).

For the prediction of the immune signature of TH1 cells, again *inflamed (tumor, CD8)* was kept in the model (**Figure 3D**, **Supplementary Table 4**). For the prediction of the immune signature of TH17 cells, *inflamed (tumor, CD4)* was kept in the model (**Figure 3F**, **Supplementary Table 5**). Both models were significantly improved by adding the TIPs (TH1 cells:  $p = 2.8e-08$ , TH17 cells:  $p = 0.0002$ ).

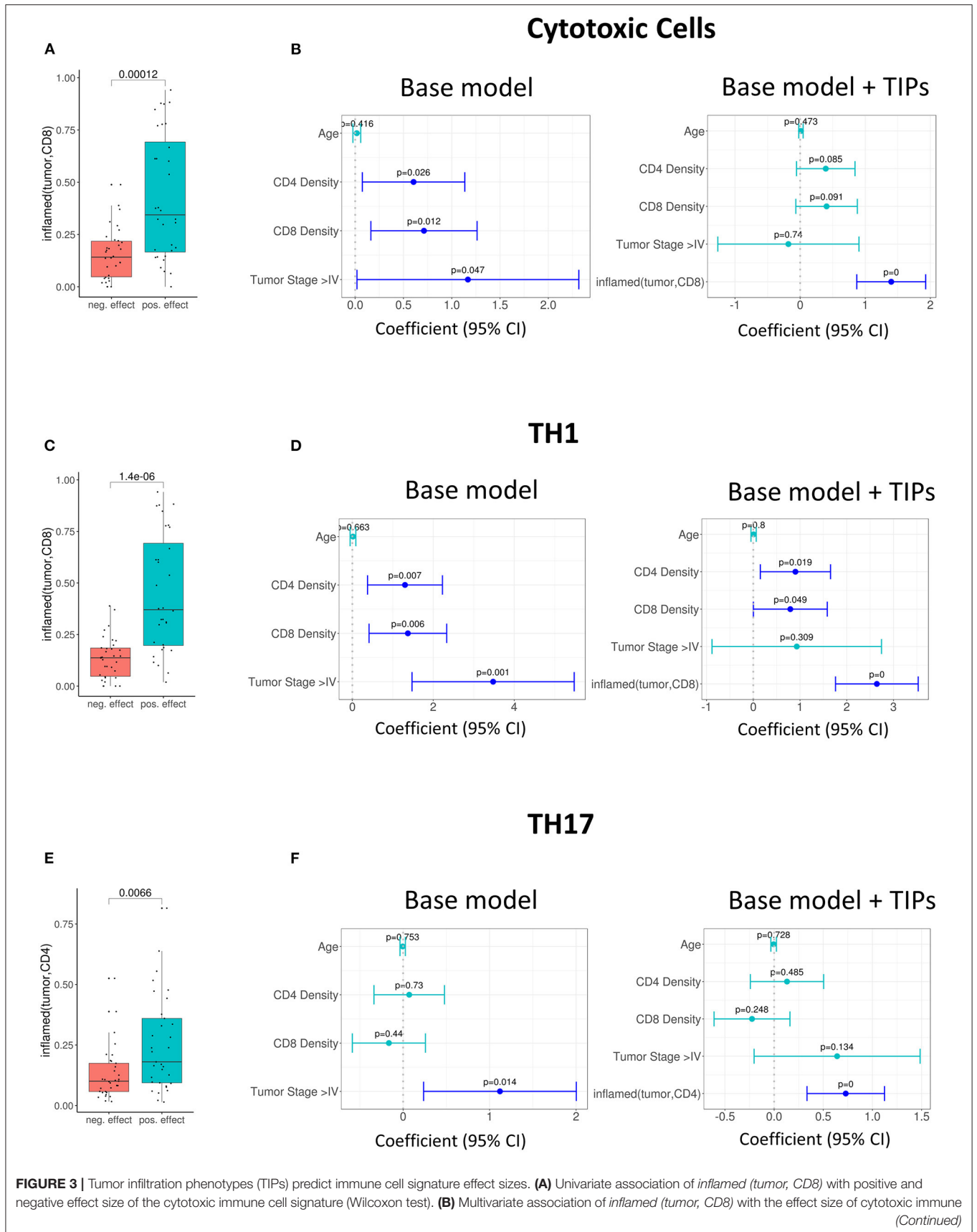
### Tumor Infiltration Phenotypes Are Associated With Mutations on the Single-Gene and Pathway Level

We investigated the association of common mutations in CRC with TIPs. For this purpose, we implemented a Lasso regularized



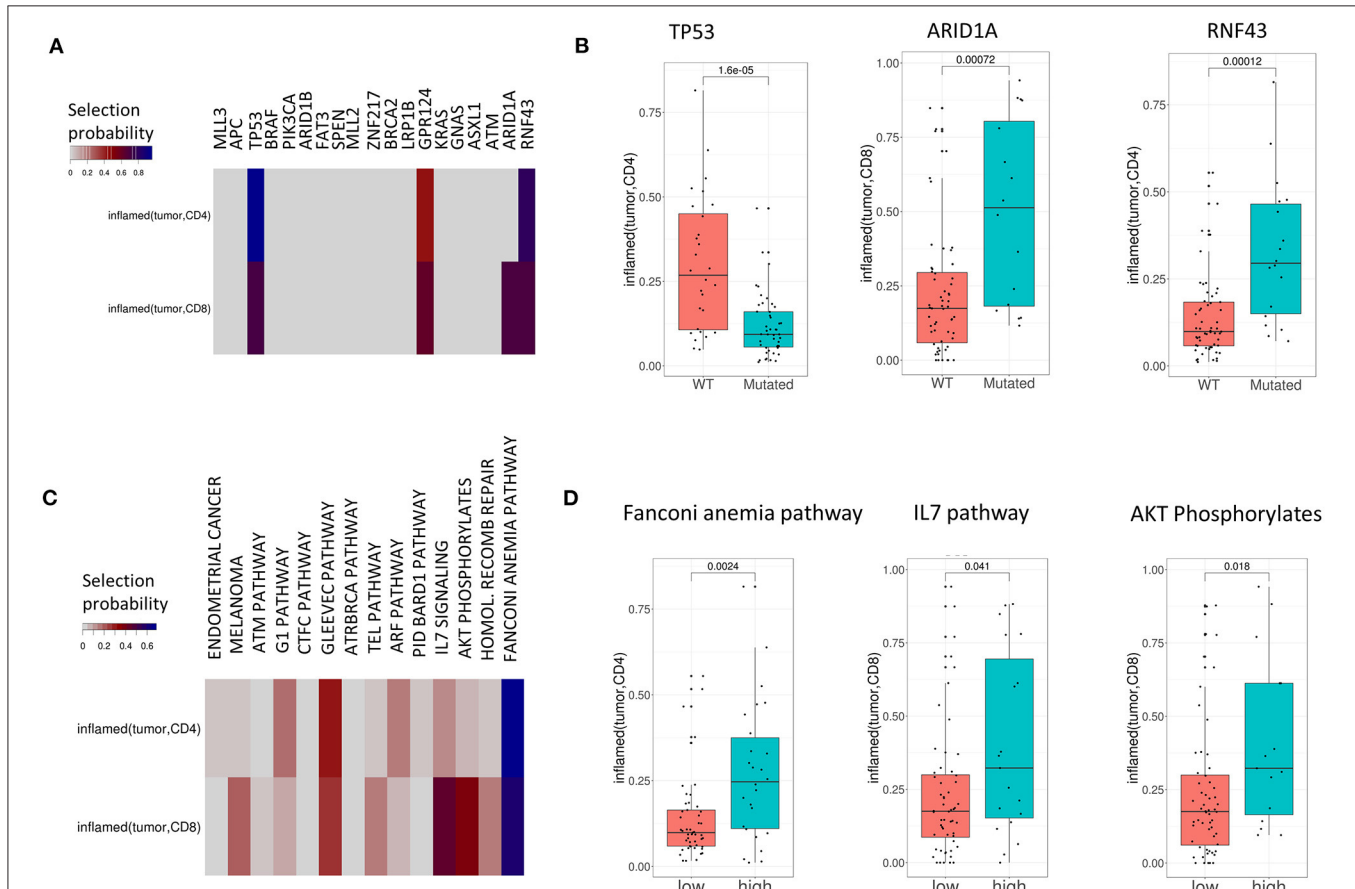
regression model that explains each TIP (*inflamed* and *excluded*) using the patients' mutation profiles as covariates (Figure 4). The analysis was performed both on the single-gene level (Figure 4A) and the pathway level (Figure 4C) using single-gene mutations that occurred in at least 20% of patient samples and 14 selected

pathways with at least 40% overlap with the set of mutated genes, respectively (see Methods and Supplementary Table 6). In order to robustly select pathways and mutations that are associated with TIPs, we used a stability selection approach for the Lasso model (35, 40), using the selection probability as an





**FIGURE 3** | cell signature. *Inflamed (tumor, CD8)* improves the base model significantly [likelihood ratio (LR) test = 5.01e-0.7]. **(C)** Univariate association of *inflamed (tumor, CD8)* with positive and negative effect size of T-helper Cell 1 immune cell signature. **(D)** Multivariate association of *inflamed (tumor, CD8)* with the effect size of T-helper Cell 1 immune cell signature. The base model is significantly improved [likelihood ratio (LR) test = 2.8e-08]. **(E)** Univariate association of *inflamed (tumor, CD4)* with positive and negative effect size of T helper cell 17 immune cell signature (Wilcoxon test). **(F)** Multivariate association of *inflamed (tumor, CD4)* with the effect size of T-helper Cell 17 immune cell signature [likelihood ratio (LR) test = 0.0002]. In all forest plots, the X-axis represents the regression coefficients and 95% intervals (whiskers). The dashed vertical line represents a regression coefficient of 0. Significant covariates are indicated in blue, non-significant covariates are indicated in turquoise.



**FIGURE 4** | TIPs are associated with mutations both on the single-gene and pathway level. **(A)** Associations between mutations and TIPs by Lasso stability selection. The coefficients represent the selection probability in 200 runs of the Lasso model. TP53, ARID1A, and RNF43 showed the strongest associations with a TIP. **(B)** Associations of mutated genes with TIPs (Wilcoxon test). **(C)** Associations between mutations in signaling pathways (pathway mutation scores) and TIPs by Lasso stability selection. The coefficients represent the selection probability in 200 runs of the Lasso model. The Fanconi Anemia pathway, the IL7 signaling pathway and the AKT phosphorylates pathway showed the strongest associations with TIPs. **(D)** Association of high and low pathway mutation scores with TIPs (Wilcoxon test). Pathway mutation scores are split into low and high at the median.

indicator of the association strength of a gene or a pathway with a TIP.

On the single-gene level, four out of 19 genes were selected in at least 40% of the runs (TP53, GPR124, ARID1A, and RNF43) reflecting the importance of these genes in explaining different TIPs. The tumor suppressor gene p53 ( $p = 1.6e-05$ , Wilcoxon Test, **Figure 4B**) and the kinases ARID1A ( $p = 0.0007$ , Wilcoxon Test, **Figure 4B**) and RNF43 ( $p = 0.00012$ , Wilcoxon Test, **Figure 4B**) showed the strongest single-gene associations with a TIP in terms of selection probability (**Figure 4A**), indicating

these genes to be relevant factors for the TIPs *inflamed (tumor, CD4)* (TP53, RNF43) and *inflamed (tumor, CD8)* (ARID1A), respectively. On the pathway level, 12 of the 14 pathways were found to be associated with at least one TIP. Here, mutations in the *Fanconi Anemia* pathway ( $p = 0.0024$ , Wilcoxon Test, **Figure 4D**), the *IL-7 signaling* pathway ( $p = 0.041$ , Wilcoxon Test, **Figure 4D**) and the *AKT* pathway ( $p = 0.018$ , Wilcoxon Test, **Figure 4D**) showed the strongest associations with a TIP [*inflamed (tumor, CD4)*: *Fanconi Anemia* pathway, *inflamed (tumor, CD8)*: *IL-7 signaling* pathway, *AKT* pathway].

## Tumor Infiltration Phenotypes Are Associated With Gene Expression Changes in Immune-Related Pathways

We hypothesized that TIPs evoke distinct gene regulatory programs in lymphocytes and tumor cells, respectively. Therefore, we were interested in detecting differences in gene expression that cannot be explained by variation in cell type composition, but which result from changes in pathway activity in either lymphocytes or tumor cells. For this purpose, we developed an approach for the deconvolution of gene expression, separating variation in gene expression due to cell proportions from changes in gene expression resulting from distinct regulatory programs (Figure 5A, Methods). Afterwards, we determined the association of TIPs to differentially expressed genes by linear regression analysis, with the TIP as the response and the gene expression values as covariates. Genes with regression coefficients significantly different from zero ( $p < 0.05$  after Bonferroni multiple testing correction) were added to the respective TIP's gene signature.

We found 303 genes to be significantly associated with *inflamed (tumor, CD8)* (Figure 5B, Supplementary Table 7), and 87 genes to be associated with *inflamed (tumor, CD4)* (Supplementary Figure 2, Supplementary Table 8).

Gene ontology (GO) term enrichment analysis on the upregulated genes for the TIP *inflamed (tumor, CD8)* revealed terms associated with the immune system (including immune response, leukocyte activation, immune effector process) (Figure 5B). The downregulated genes were found to be associated to dendrite morphogenesis regulation and transcription initiation. Among the significantly downregulated genes we identified the DNA mismatch repair protein MLH1 (41), the genes CRCP and ZXDA with roles in DNA transcription regulation and BBS10 with a role in protein folding (42, 43). The most significant upregulated genes included histocompatibility antigens like HLA-DMA, HLA-DPA1, HLA-DRA, HLA-DPB1, HLA-DMB, and genes with a role in immune response (CD74, TCIRG1, RAC2, ITGAE).

GO terms in the gene signature for *inflamed (tumor, CD4)* were enriched for neutrophil migration, neutrophil chemotaxis and regulation of dendritic spine (Supplementary Figure 2).

## Genetic Signatures Predict Tumor Infiltration Phenotypes

Gene signatures allow extending the definition of TIPs from the tissue to the molecular level. In order to test if the derived TIP-specific gene signatures can also be used to predict TIPs in new datasets with missing IHC image data, we obtained two genomic datasets from CRC patients, TCGA-COAD ( $n = 298$ ) and TCGA-READ ( $n = 98$ ), using the Cancer Genome Atlas (TCGA).

First, we trained a support vector machine on the original CRC ( $n = 80$ ) dataset with the TIPs as response variables and the genes of the corresponding gene signature as features (Figure 5C), learning the association between gene expression and TIPs. Afterwards we applied this model to predict TIPs in the genomic datasets from TCGA. We found the predicted TIP *inflamed (tumor, CD8)* to be significantly higher in MSI-High

patients than in MSI-Low or MSS (Microsatellite stable) patients as defined in TCGA (Figure 5D,  $p = 3.7e-16$ ). Further, we found the TIP *inflamed (tumor, CD8)* to be significantly higher in right side tumors as compared to left side tumors (Figure 5D,  $p = 1.1e-06$ ), confirming the hypothesis of right side tumors having a stronger immune infiltrated tumor microenvironment (44).

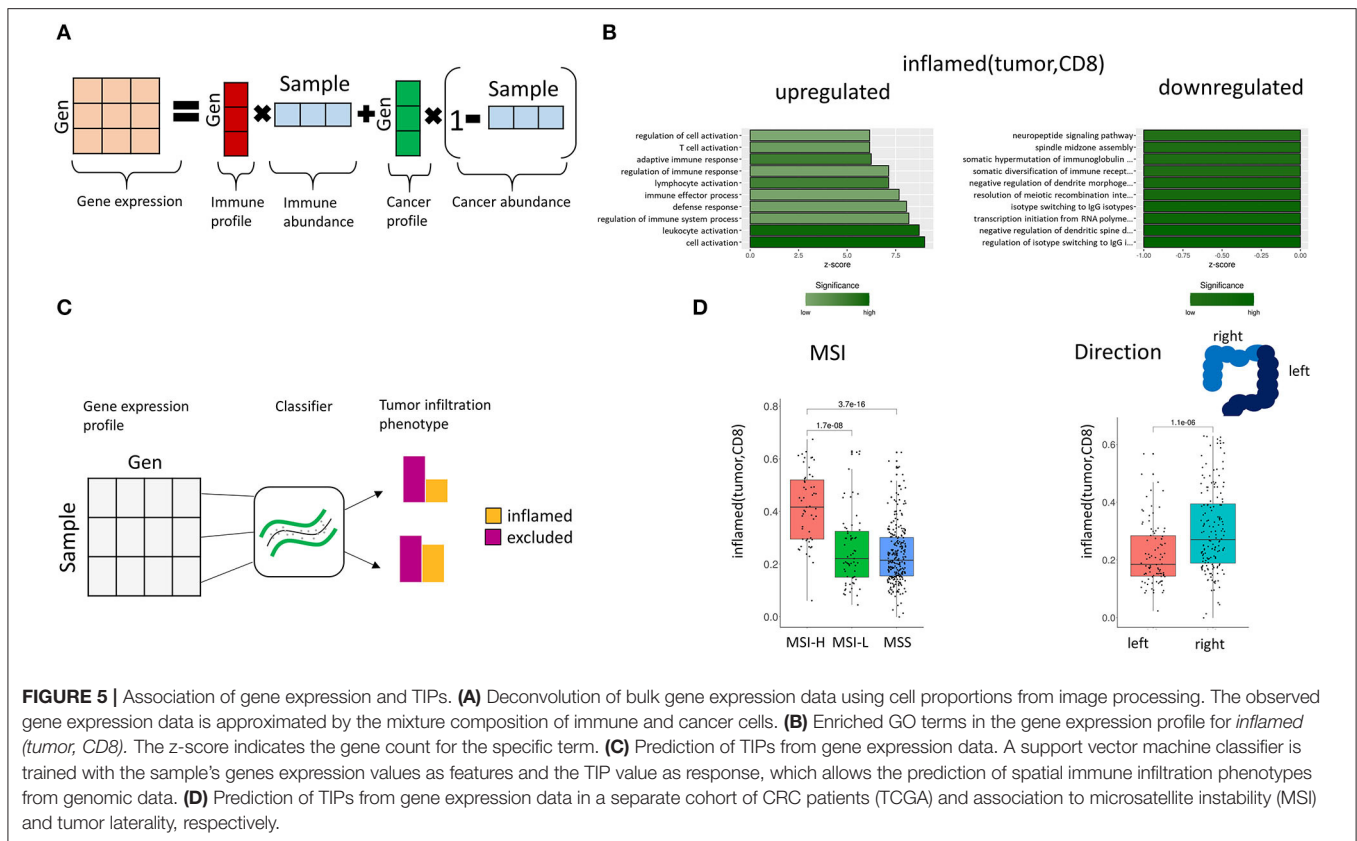
## DISCUSSION

The quantification of spatial tumor immune infiltration has great potential to form patient enrichment strategies for immunotherapy, or to help to better understand the mode of action of a drug in early clinical trials. For instance, the co-localization of immune cells and tumor cells has been proven to be connected to beneficial survival in many tumor types (14, 15). Here, we have developed a method that is based on the three basic immune infiltration patterns *inflamed/excluded/deserted* but, in contrast to other approaches, is independent of manual annotations and thus allows the automated, quantitative identification from immunohistochemically stained samples. In comparison to other studies that used spatial statistics to infer spatial patterns of immune and tumor cells (14, 45), our method considers the intra-tumor heterogeneity of immune infiltration. This is achieved by dividing the tissue into tiles and applying Ripley's L statistic in order to classify the tiles according to their local immune infiltration pattern. The proportion of the local immune infiltration patterns in the sample serves as our measurement of the global tumor infiltration phenotype (TIP).

We have demonstrated that TIPs are associated with the MSI and TMB status of a patient. While it has been shown previously that MSI and high TMB are associated with a higher immune cell density (46), we could show that they are also associated with an increased co-localization between CD8+ T cells and tumor cells. Moreover, TIPs were shown to be independent from other covariates like cell density, age and tumor stage and thus served as an additional factor for the prediction of the MSI and TMB status, respectively.

Notably, the effect size of the immune signature for Cytotoxic T cells was correlated with increased co-localization between tumor and CD8+ T cells. This highlights the great potential to combine spatial infiltration phenotypes with phenotypes derived from gene expression, like the subpopulation of cytotoxic cells in order to predict the outcome of immunotherapy. Nevertheless, the prognostic and predictive relevance of phenotype combinations must be determined in future studies.

Furthermore, we found TIPs to be associated with common mutations in CRC. It has been shown that immune infiltration is triggered by somatic mutations. Whereas, on the one hand mutations can induce immune escape mechanisms in the tumor cells (47), they can on the other hand also form neoantigens that are detected by the immune system (48). We showed that mutations in p53 resulted in reduced co-localization between tumor and CD8+ T cells whereas mutations in ARID1A and RNF43 had the opposite effect. A reason may be that p53 dysfunction leads to immunosuppression and immune evasion



**FIGURE 5 |** Association of gene expression and TIPs. **(A)** Deconvolution of bulk gene expression data using cell proportions from image processing. The observed gene expression data is approximated by the mixture composition of immune and cancer cells. **(B)** Enriched GO terms in the gene expression profile for *inflamed (tumor, CD8)*. The z-score indicates the gene count for the specific term. **(C)** Prediction of TIPs from gene expression data. A support vector machine classifier is trained with the sample's genes expression values as features and the TIP value as response, which allows the prediction of spatial immune infiltration phenotypes from genomic data. **(D)** Prediction of TIPs from gene expression data in a separate cohort of CRC patients (TCGA) and association to microsatellite instability (MSI) and tumor laterality, respectively.

(49) whereas mutations in ARID1A and RNF43 might act as neoantigens. Notably, mutations in the Fanconi Anemia pathway were associated with an increased co-localization between tumor cells and CD4 cells. As the Fanconi pathway is a DNA mismatch repair pathway (50), perturbation of this pathway causes an accumulation of replication errors, making cancer cells more easily recognizable by cytotoxic cells (51). Moreover, mutations in the AKT pathway and the IL7 signaling pathway were connected with TIPs. The AKT pathway is a cell signaling pathway. It is well-known that activation of AKT is a mechanism of tumor immune evasion (52), which is corresponding to the observation that high mutation rates in this pathway were in our analysis connected with more co-localization between tumor cells and CD8 cells. Interleukin-7 is an important cytokine that stimulates B and T cell development. However, it has been shown that in tumor cells IL-7 is associated with increased aggressiveness of solid tumors and enlarged metastasis rate (53, 54). Although the association of high mutation rates in these pathways with TIPs is less obvious, it shows the potential of TIPs for the characterization of the immune response when combined with genomic data.

Following this line of thought, we integrated spatial immune infiltration patterns, as represented by the TIPs, with gene expression data. Using a gene expression deconvolution scheme in combination with a regression model, we were able to find specific gene signatures for TIPs. We found upregulated genes in the signature of *inflamed (tumor, CD8)* to be connected to

the immune system whereas downregulated genes were shown to have a role DNA in mismatch repair. This supports the hypothesis that errors in DNA replication are the main driver for the immune answer in CRC.

Interestingly, these gene signatures can be seen as a representation of TIPs on the genomic level and allow the determination of TIPs also in data sources for which no IHC imaging data is available. We have demonstrated this approach in a CRC dataset from TCGA and showed that the predicted TIPs were associated with MSI-high tumors and the laterality of the tumor. The definition of TIPs on the genomic level, enable their use in a much broader range of data sources, e.g., in public databases or phase 3 clinical trials for which typically no IHC stainings are available.

In conclusion, we have demonstrated how tumor infiltration phenotypes can be quantified in CRC samples using a computational method that accounts for the high variability of spatial immune infiltration across the sample. The computational tumor infiltration phenotypes do not only have the potential to serve as biomarkers in immunotherapy but also enable a more detailed exploration of the immune response in CRC. Although immunotherapy has been successful in CRC in the last years, response rates are variable. In order to find the right treatment for every patient, a detailed characterization of individual tumors is necessary. Despite the established classifications for CRC, e.g., by consensus molecular subtypes or the MSI status, tumor infiltration phenotypes

offer the potential for an even deeper characterization of CRC patients.

In the future, it would be of great interest to include more cell types into the analysis e.g., myeloid cells or fibroblasts to capture additional factors in the tumor microenvironment. Moreover, it remains to be shown whether TIPs can serve as prognostic or predictive biomarkers by testing their association to endpoints such as overall survival or treatment response.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

Samples used in this study were provided by Avaden Biosciences, Seattle WA 98104, USA and IndividuMed GmbH, 20251 Hamburg, Germany. Written agreements for all relevant samples were signed, guaranteeing that appropriate ethics approvals were obtained.

## REFERENCES

- Geospatial immune variability illuminates differential evolution of lung adenocarcinoma | Nature Medicine. Available online at: <https://www.nature.com/articles/s41591-020-0900-x>
- George AP, Kuzel TM, Zhang Y, Zhang B. The discovery of biomarkers in cancer immunotherapy. *Comput Struct Biotechnol J*. (2019) 17:484–97. doi: 10.1016/j.csbj.2019.03.015
- Boumber Y. Tumor mutational burden (TMB) as a biomarker of response to immunotherapy in small cell lung cancer. *J Thorac Dis*. (2018) 10:4689–93. doi: 10.21037/jtd.2018.07.120
- Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* (2019) 30:44–56. doi: 10.1093/annonc/mdy495
- Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I–III colon cancer. *Cancer Immunol Immunother.* (2018) 68:433–42. doi: 10.1007/s00262-018-2289-7
- Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin F, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sic Trans Med.* (2012) 4:157ra143. doi: 10.1126/scitranslmed.3004330
- Murciano-Goroff YR, Warner AB, Wolchok JD. The future of cancer immunotherapy: microenvironment-targeting combinations. *Cell Res.* (2020) 30:507–19. doi: 10.1038/s41422-020-0337-2
- AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med.* (2020) 26:1054–62. doi: 10.1038/s41591-020-0900-x
- Trujillo JA, Sweis RF, Bao R, Luke JJ. T cell-inflamed versus non-T cell-inflamed tumors: a conceptual framework for cancer immunotherapy drug development and combination therapy selection. *Cancer Immunol Res.* (2018) 6:990–1000. doi: 10.1158/2326-6066.CIR-18-0277
- Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* (2016) 17:174. doi: 10.1186/s13059-016-1028-7

## AUTHOR CONTRIBUTIONS

FS and KK initiated the project. HF, FS, KK, AT, and NZ conceived the idea of computational tumor infiltration phenotypes. HF implemented the analysis. FS and AT provided statistical guidance. NZ and KK provided biological interpretation. HF and FS wrote the manuscript. HF, FS, KK, and AT revised the manuscript. All authors read and approved the submitted version.

## ACKNOWLEDGMENTS

The authors thank Sabine Moosmann, Eldad Klaiman and Dr. Chia-Huey Ooi for their support with the IF panel and immune signatures. The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.552331/full#supplementary-material>

- Feldmeyer L, Hudgens CW, Ray-Lyons G, Nagarajan P, Aung PP, Curry JL, et al. Density, distribution, and composition of immune infiltrates correlate with survival in merkel cell carcinoma. *Clin Cancer Res.* (2016) 22:5553–63. doi: 10.1158/1078-0432.CCR-16-0392
- Koelzer VH, Sirinukunwattana K, Rittscher J, Mertz KD. Precision immunoprofiling by image analysis and artificial intelligence. *Virchows Arch.* (2019) 474:511–22. doi: 10.1007/s00428-018-2485-z
- Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature.* (2017) 541:321–30. doi: 10.1038/nature21349
- Carstens JL, Correa de Sampaio P, Yang D, Barua S, Wang H, Rao A, et al. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun.* (2017) 8:15095. doi: 10.1038/ncomms15095
- Nawaz S, Heindl A, Koelble K, Yuan Y. Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Modern Pathol.* (2015) 28:766–77. doi: 10.1038/modpathol.2015.37
- Ripley BD. The second-order analysis of stationary point processes. *J Appl Probability.* (1976) 13:255–66. doi: 10.1017/S0021900200094328
- Arnold T, Tibshirani R. Efficient implementations of the generalized lasso dual path algorithm. *arXiv:1405.3222 [cs, stat]* (2014).
- Tibshirani RJ, Taylor J. The solution path of the generalized lasso. *Ann Statist.* (2011) 39:1335–71. doi: 10.1214/11-AOS878
- Zhang W, Hubbard A, Jones T, Racolta A, Bhaumik S, Cummins N, et al. Fully automated 5-plex fluorescent immunohistochemistry with tyramide signal amplification and same species antibodies. *Lab. Invest.* (2017) 97:873–85. doi: 10.1038/labinvest.2017.37
- Wong HL, Christie M, Gately L, Tie J, Lee B, Semira C, et al. Mismatch repair deficiency assessment by immunohistochemistry: for Lynch syndrome screening and beyond. *Future Oncol.* (2018) 14:2725–39. doi: 10.2217/fon-2018-0319
- Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol.* (2013) 31:1023–31. doi: 10.1038/nbt.2696
- Analysis of Spatial Organization of Suppressive Myeloid Cells and Effector T Cells in Colorectal Cancer—A Potential Tool for Discovering Prognostic

- Biomarkers in Clinical Research. Available online at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7658632/>
23. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biology*. (2015) 16:64. doi: 10.1186/s13059-015-0620-6
  24. Sandmann T, Kummerfeld SK, Gentleman R, Bourgon R. gCMAP: user-friendly connectivity mapping with R. *Bioinformatics*. (2014) 30:127–8. doi: 10.1093/bioinformatics/btt592
  25. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. (2012) 40:e133. doi: 10.1093/nar/gks461
  26. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. (2016) 44:e71. doi: 10.1093/nar/gkv1507
  27. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. (2000) 28:27–30. doi: 10.1093/nar/28.1.27
  28. Spatial Point Patterns: Methodology and Applications with R - CRC Press Book. Available online at: <https://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200>
  29. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. (2010) 11:R106. doi: 10.1186/gb-2010-11-10-r106
  30. Baddeley A, Rubak E, Turner R. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, (2015). doi: 10.1201/b19708
  31. Baddeley A, Turner R. spatstat: an R package for analyzing spatial point patterns. *J Statistical Software*. (2005) 12:1–42. doi: 10.18637/jss.v012.i06
  32. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J*. (2010) 52:70–84. doi: 10.1002/bimj.200900028
  33. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag, (2016). doi: 10.1007/978-3-319-24277-4\_9
  34. Kuznetsova A, Brockhoff PB, Christensen RHB. **ImerTest** package: tests in linear mixed effects models. *J Stat Soft*. (2017) 82:1–26. doi: 10.18637/jss.v082.i13
  35. Yuan Y, Curtis C, Caldas C, Markowitz F. A sparse regulatory network of copy-number driven gene expression reveals putative breast cancer oncogenes. *IEEE/ACM Trans Comput Biol Bioinform*. (2012) 9:947–54. doi: 10.1109/TCBB.2011.105
  36. Fu A, Narasimhan B, Boyd S. CVXR: an R package for disciplined convex optimization. *arXiv:1711.07582 [stat]*. (2017).
  37. Alexa ARJ. *topGO: Enrichment Analysis for Gene Ontology*. in. (2019).
  38. Dimitriadou, Evgenia, Hornik, Kurt, Leisch, Friedrich, Meyer, David & Weingessel, Andreas. e1071: Misc Functions of the Department of Statistics (e1071).
  39. Riviere P, Goodman AM, Okamura R, Barkauskas DA, Whitchurch TJ, Lee S, et al. High tumor mutational burden correlates with longer survival in immunotherapy-naïve patients with diverse cancers. *Mol Cancer Ther*. (2020) 19:2139–45. doi: 10.1158/1535-7163.MCT-20-0161
  40. Yuan Y (2019). lol: Lots Of Lasso. R package version 1.32.0.
  41. Pal T, Permeth-Wey J, Sellers TA. A review of the clinical relevance of mismatch-repair deficiency in ovarian cancer. *Cancer*. (2008) 113:733–42. doi: 10.1002/cncr.23601
  42. Seo S, Baye LM, Schulz NP, Beck JS, Zhang Q, Slusarski DC, et al. BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proc Natl Acad Sci USA*. (2010) 107:1488–93. doi: 10.1073/pnas.0910268107
  43. Nowsheen S, Aziz K, Luo K, Deng M, Qin B, Yuan J, et al. ZNF506-dependent positive feedback loop regulates H2AX signaling after DNA damage. *Nat Commun*. (2018) 9:1–11. doi: 10.1038/s41467-018-05161-0
  44. Zhang L, Zhao Y, Dai Y, Cheng, J.-N., Gong Z, et al. Immune landscape of colorectal cancer tumor microenvironment from different primary tumor location. *Front Immunol*. (2018) 9:1578. doi: 10.3389/fimmu.2018.01578
  45. Schwen LO, Andersson E, Korski K, Weiss N, Haase S, Gaire F, et al. Data-driven discovery of immune contexture biomarkers. *Front Oncol*. (2018) 8:627. doi: 10.3389/fonc.2018.00627
  46. Maby P, Tougeron D, Hamieh M, Mlecnik B, Kora H, Bindea G, et al. Correlation between density of CD8+ T-cell infiltrate in microsatellite unstable colorectal cancers and frameshift mutations: a rationale for personalized immunotherapy. *Cancer Res*. (2015) 75:3446–55. doi: 10.1158/0008-5472.CAN-14-3051
  47. Quigley D, Silwal-Pandit L, Dannenfels R, Langerød A, Volla HKM, Vaske C, et al. Lymphocyte invasion in IC10/Basal-Like Breast Tumors Is Associated With Wild-type TP53. *Mol Cancer Res*. (2015) 13:493–501. doi: 10.1158/1541-7786.MCR-14-0387
  48. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. (2015) 348:69–74. doi: 10.1126/science.aaa4971
  49. Cui Y, Guo G. Immunomodulatory function of the tumor suppressor p53 in host immune response and the tumor microenvironment. *Int J Mol Sci*. (2016) 17:1942. doi: 10.3390/ijms17111942
  50. Walden H, Deans AJ. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys*. (2014) 43:257–78. doi: 10.1146/annurev-biophys-051013-022737
  51. Parkes EE, Walker SM, Taggart LE, McCabe N, Knight LA, Wilkinson R, et al. Activation of STING-dependent innate immune signaling by S-phase-specific DNA damage in breast cancer. *J Natl Cancer Inst*. (2016) 109:1. doi: 10.1093/jnci/djw199
  52. Noh KH, Kang TH, Kim JH, Pai SI, Lin KY, Hung F, et al. Activation of Akt as a mechanism for tumor immune evasion. *Molecular Therapy*. (2009) 17:439–47. doi: 10.1038/mt.2008.255
  53. Krzystek-Korpacka M, Zawadzki M, Neubauer K, Bednarz-Misa I, Górska S, Wiśniewski J, et al. Elevated systemic interleukin-7 in patients with colorectal cancer and individuals at high risk of cancer: association with lymph node involvement and tumor location in the right colon. *Cancer Immunol Immunother*. (2017) 66:171–9. doi: 10.1007/s00262-016-1933-3
  54. Al-Rawi MAA, Rmali K, Watkins G, Mansel RE, Jiang WG. Aberrant expression of interleukin-7 (IL-7) and its signalling complex in human breast cancer. *Eur J Cancer*. (2004) 40:494–502. doi: 10.1016/j.ejca.2003.10.016

**Conflict of Interest:** HE, NZ, KK, and FS are employees of F. Hoffmann-La Roche Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Failmezger, Zwigg, Tresch, Korski and Schmich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.