Received: 5 January 2021 | Accepted: 18 August 2022

DOI: 10.1111/biom.13774

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Bayesian hierarchical quantile regression with application to characterizing the immune architecture of lung cancer

Priyam Das[1] | **Christine B. Peterson**[2] | **Yang Ni**[3] | **Alexandre Reuben**[4] | **Jiexin Zhang**[5] | **Jianjun Zhang**[4] | **Kim-Anh Do**[2] | **Veerabhadran Baladandayuthapani**[6]

[1]Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

[2]Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[3]Department of Statistics, Texas A&M University, College Station, Texas, USA

[4]Department of Thoracic Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[5]Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[6]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Veerabhadran Baladandayuthapani, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109, USA.
Email: veerab@umich.edu

Funding information
Center for Strategic Scientific Initiatives, National Cancer Institute, Grant/Award Numbers: R01-CA160736, R01CA244845-01A1, R21-CA220299, P30; Division of Mathematical Sciences, Grant/Award Numbers: DMS-1463233, DMS-1918851; Cancer Prevention and Research Institute of Texas (CIPRIT), Grant/Award Numbers: RP150521, RP160693

Abstract

The successful development and implementation of precision immuno-oncology therapies requires a deeper understanding of the immune architecture at a patient level. T-cell receptor (TCR) repertoire sequencing is a relatively new technology that enables monitoring of T-cells, a subset of immune cells that play a central role in modulating immune response. These immunologic relationships are complex and are governed by various distributional aspects of an individual patient's tumor profile. We propose Bayesian QUANTIle regression for hierarchical COvariates (QUANTICO) that allows simultaneous modeling of hierarchical relationships between multilevel covariates, conducts explicit variable selection, estimates quantile and patient-specific coefficient effects, to induce individualized inference. We show QUANTICO outperforms existing approaches in multiple simulation scenarios. We demonstrate the utility of QUANTICO to investigate the effect of TCR variables on immune response in a cohort of lung cancer patients. At population level, our analyses reveal the mechanistic role of T-cell proportion on the immune cell abundance, with tumor mutation burden as an important factor modulating this relationship. At a patient level, we find several outlier patients based on their quantile-specific coefficient functions, who have higher mutational rates and different smoking history.

KEYWORDS
Bayesian quantile regression, non-small-cell lung cancer, T-cell receptor repertoire, variable selection, varying sparsity regression

1 | INTRODUCTION

Immunotherapy is a class of cancer treatments that fosters the patient's own immune system to fight cancer (Waldman et al., 2020). Although immunotherapies represent a major breakthrough in cancer treatment, offering immense clinical benefit to some patients with a lower toxicity burden than chemotherapy, resistance to immunotherapy remains a major challenge (Walsh & Soo, 2020). Tumors use various strategies to protect themselves from antitumor immunity which might vary across patients. Antitumor immune responses might also be mediated by several different mechanisms, driven by patient-specific immune architecture. Cancer immunotherapy therefore needs to be personalized to recognize patient-specific rate-limiting steps and employ strategies for overcoming these hurdles (Kakimi et al., 2017).

Identifying patient-specific influences on immune response requires integrating measures of immune activity and characterizing their dependence on upstream factors. Here, we consider a setting where the response variable Y is a continuous measure of immune activation, such as the abundance of CD8+ T-cells, which play a key role in directly killing cancer cells. The abundance of these cells is driven by "Level 1" covariates $T_1, \ldots, T_p$, which represent measures of upstream immune activity. Here we take these Level 1 covariates to be features derived from T-cell receptor (TCR) sequencing, which capture activity of the adaptive immune system. The TCR repertoire depends, in turn on "Level 2" influences further upstream, such as antigens that the T-cells have been exposed to. This includes tumor-specific antigens, which are abnormal proteins produced by tumor cells due to DNA mutations in the cell (see Figure 1A). We consider mutational variables $M_1, \ldots, M_g$ as (hierarchical) Level 2 covariates in our model. Our goal is to develop a hierarchical modeling approach providing insights into the mechanistic relationships among these variables wherein the measures of immune activity may depend on the upstream factors in a complex nonlinear fashion.

We now briefly review prior work addressing the challenge of flexible regression modeling, which lays the foundation for the proposed approach. In order to estimate the subject-specific effect of variables on the outcome variable, Hastie and Tibshirani (1993) proposed the varying coefficient model (VCM). Since then, several variations of the VCM have been proposed (Fan & Zhang, 1999; Park et al., 2013), including approaches that incorporate shrinkage in estimation (Wang & Xia, 2009). However, existing VCM methods do not enable explicit multilevel variable selection, which is crucial in settings such as ours where there are a large number of explanatory variables within a hierarchy. Although VCMs allow more flexible coefficients than traditional regression models, they are still focused on estimation of the mean of the response variable. If one is interested in obtaining a comprehensive picture of the effect of the predictors on the response variable, mean regression might be insufficient. For example, if the dependent variable is multimodal or skewed, estimating the mean effect might be misleading. In such scenarios, median or more generally, quantile regression may be more appropriate (Koenkar & Bassett, 1978). Specifically, if the interest lies specifically at higher or lower quantiles of the response variable, quantile regression is more suitable.

Over the last couple of decades, methodological developments in quantile regression have been proposed in both the classical and Bayesian frameworks (Das & Ghosal, 2018; Kottas & Gelfand, 2001; Reich, 2012; Yu & Moyeed, 2001). Several articles emerged in the literature integrating quantile regression with VCMs (Honda, 2004; Kim, 2007; Tang & Wang, 2005). In Bayesian settings, the literature on VCMs is relatively sparse (Biller & Fahrmeir, 2001), and these proposals do not address quantile regression. Recently, Ni et al. (2018) proposed a VCM incorporating variable selection in the Bayesian framework, which was applied to characterize the relationship of cancer patient outcomes to proteomics and genomics data. Although there are approaches for Bayesian variable selection in multilevel models that assume linear covariate effects (Koslovsky et al., 2020; Stingo et al., 2013), to the best of our knowledge, no prior work has considered variable selection in VCM for quantile regression.

In order to understand the subject-specific effect of hierarchically structured covariates on the outcome variable we propose Bayesian QUANTIle regression for hierarchical COvariates (QUANTICO), where the regression coefficients are allowed to differ across patients for any given quantile level of the outcome variable. Among the hierarchically structured (multilevel) covariates, we consider the Level 1 covariates have a direct effect on the response variable, modulated by Level 2 covariates. Since we expect the covariate effects to be heterogeneous across patients, we want to perform individualized inference as well as allowing variable selection for both Level 1 and Level 2 covariates. As we also expect the effects to be different at different parts of the distribution of the response variable, we model across the quantiles, rather than only considering fixed moments (e.g., mean). This provides a richer, broader, and more flexible exploration of the relationship structure. We show an illustration of the proposed model in Figure 1B with n subjects, p Level 1 and g Level 2 covariates. As shown in the figure, the selection of Level 1 and 2 covariates is allowed to vary across quantiles. Due

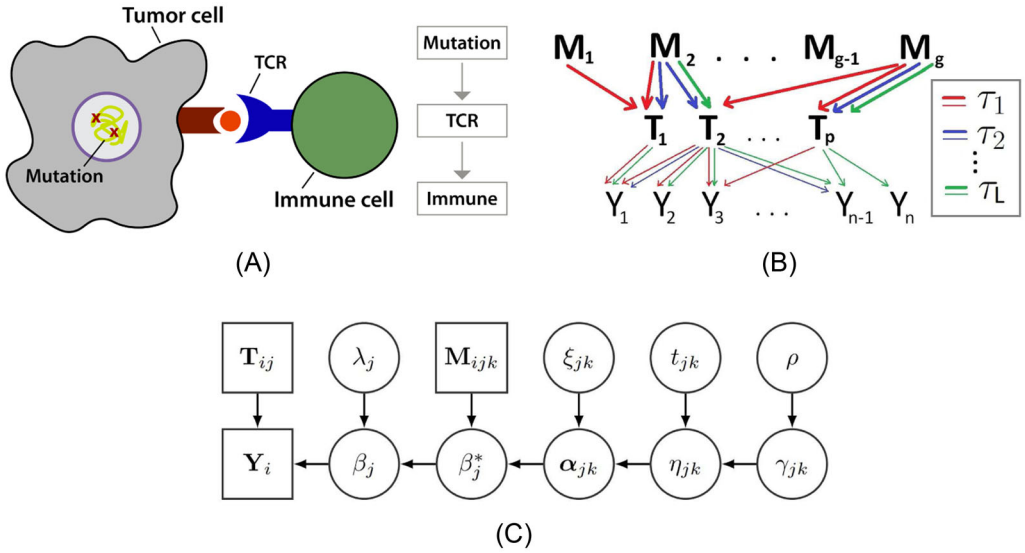

FIGURE 1 (A) Illustration of how mutation within tumor cells and T-cell receptor (TCR) repertoire impact immune cells. (B) Illustration of the proposed model for a scenario with n patients, g mutation variables $M_1, \dots, M_g$, and p TCR variables $T_1, \dots, T_p$. The response variable is denoted $Y_1, \dots, Y_n$. Different colored lines describe the estimated dependency structure at different quantile levels denoted by $(\tau_1, \tau_2, \dots \tau_L)$. (C) Directed acyclic graph (DAG) of the QUANTICO model. Parameters are shown in circles and the observed data are shown in boxes

to the presence of two levels of covariates and due to the fact that the effect of Level 2 covariates is induced on the output variable via its effect on Level 1 covariates, we call it a hierarchical model.

The rest of the paper is organized as follows: Section 2 describes the QUANTICO modeling framework, with a discussion of the priors in Section 3. We then describe the computational algorithm for performing posterior inference (Section 4), and provide a simulation study comparing the performance of the proposed method with existing alternatives (Section 5). In Section 6, we apply QUANTICO to characterize the relationship of the CD8 immune marker with TCR and mutation variables for lung cancer patients. We conclude with a brief discussion and possible extensions of our methodology in Section 7.

2 | QUANTICO MODEL

In Section 2.1, we introduce quantile regression in a VCM setting with two levels of covariates. In Section 2.2, we describe variable selection procedures on Level 1 and Level 2 covariates. Section 2.3 summarizes the likelihood constructions.

2.1 | Varying sparsity quantile regression model

Suppose there are n subjects whose response variables (immune marker values) are denoted by $Y = (Y_1, \dots, Y_n)$ and p Level 1 covariates (TCR variables) $\mathbf{T} = (T_1, \dots, T_p)$. For the ith subject, the set of Level 1 covariates are denoted as $\mathbf{T}_i = (T_{i1}, \dots, T_{ip})$ for $i = 1, \dots, n$. We assume a linear relationship between the response variable Y and the Level 1 covariates $\mathbf{T}$. Now, for the jth Level 1 covariate T_j, consider a set of q_j Level 2 covariates (mutation variables) $\mathbf{M}_j = (M_{j1}, \dots, M_{jq_j})$ for $j = 1, \dots, p$. A Level 2 covariate induces its effect on the response variable Y through its effect on the jth Level 1 covariate, that is, T_j for $j = 1, \dots, p$. Note that the same Level 2 covariate may influence multiple Level 1 covariates, and that the number q_j of Level 2 covariates can be different for each Level 1 covariate. However, in the particular case study considered in this paper, the same set of Level 2 covariates (mutation variables) are considered for all Level 1 covariates (TCR variables). For the ith subject, the set of Level 2 covariates is denoted by $\mathbf{M}_{ij} = (M_{ij1}, \dots, M_{ijq_j})$ for $i = 1, \dots, n$, $j = 1, \dots, p$. The total number of possible effects for the M variables which can be identified as a part of the coefficient of the T variables is given by $\sum_{j=1}^{p} q_j$.

Instead of estimating the effect of T and M on the mean of the response variable, our interest lies in estimating their relationship for different quantile levels. To obtain the quantile-specific estimates, we assume the effect of the covariates on Y to be dependent on the quantile level $\tau (0 < \tau < 1)$. Let

$$Q_Y(\tau | T, M) = \inf\{q : P(Y \le q | \mathbf{T} = T, \mathbf{M} = M) \ge \tau\}, \quad (1)$$

denote the τth conditional quantile ($0 < \tau < 1$) of the response variable Y at $\mathbf{T} = T, \mathbf{M} = M$. The relation between the response and the covariates at the τth quantile is given by

$$Q_{Y_i}(\tau|T, M) = \beta_0(\tau, \mathbf{M}_{i0}) + \sum_{j=1}^{p} T_{ij}\beta_j(\tau, \mathbf{M}_{ij}), \quad (2)$$

where $\mathbf{M}_{i0}$ denotes the values of all distinct Level 2 covariates for the ith subject, and $\mathbf{M}_{ij}$ denotes the values of the Level 2 covariates corresponding to the jth Level 1 covariate for ith subject. This equation shows the dependence structure of the τth quantile of the dependent variable Y for the ith patient on the corresponding Level 1 (T) and Level 2 (M) covariates. Note that for each Level 1 covariate T_j we have q_j many distinct M variables, but since two different Level 1 covariates may share the same M variables, the $\sum_{j=1}^{p} q_j$ many M variables may not be distinct. Also note that both the intercept and slope terms are patient-specific (indexed by i). In order to incorporate the effect of the M variables on the dependent variable, at any given quantile level τ, all the slope terms (of T) are semiparametric functions of the M variables and likewise the intercept is a semiparametric function of the distinct M variables. The explicit structure of $\beta_j(\tau, \mathbf{M}_{ij})$ is discussed in Section 2.2.

2.2 | Varying-sparsity coefficient modeling and selection

For any given quantile level τ, we consider $\beta_j(\tau, \mathbf{M}_{ij})$ as a smooth function of $\mathbf{M}_{ij}$. The main motivation for taking $\beta_j(\tau, \mathbf{M}_{ij})$ as a smooth function of $\mathbf{M}_{ij}$ is so that the coefficients of any Level 1 covariate (T_j) for two different subjects are similar if the values of the corresponding Level 2 covariates ($\mathbf{M}_{\cdot j}$) are similar as well. Since for any Level 1 covariate, "neighboring" patients (with respect to the M variables) are expected to have a similar slope coefficient, this assumption enables borrowing of strength, increasing our power to estimate the subject-specific slope coefficients. For modeling the slope and intercept terms, we use spline functions due to their flexible construction, interpretation, and the ease of incorporating penalization.

At any given quantile τ ($0 < \tau < 1$), the slope and the intercept terms are estimated as the sum of spline functions given by

$$\beta_j(\tau, \mathbf{M}_{ij}) = \sum_{k=1}^{q_j} f_{jk}^{(\tau)}(M_{ijk}), j = 0, \ldots p, \quad (3)$$

where $\beta_0(\tau, \mathbf{M}_{i0})$ denotes the (global) intercept term, and $\beta_j(\tau, \mathbf{M}_{ij})$ for $j \geq 1$ denotes the slope term. The spline components $f_{jk}^{(\tau)}(M_{ijk}) = \mathbf{S_{ijk}}\boldsymbol{\alpha}_{jk}^{\tau}$ where $\mathbf{S_{ijk}}$ denotes the cubic B-spline bases for M_{ijk} and $\boldsymbol{\alpha}_{jk}^{\tau}$ denotes the corresponding spline coefficient. Note that we consider the intercept term to be a function of all Level 2 covariates. The number of knots for the B-spline bases is taken to be sufficiently large to capture the nonlinear features. We do not perform knot selection; rather equally spaced quantile knots on each of the Level 2 covariates are considered. Smoothing is induced via regularization and overfitting is controlled through a roughness penalty on the spline coefficients, details of which are provided in Section A of the Supporting Information.

In Equation (3), $\beta_j(\tau, \cdot)$ is modeled as a sum of a set of smooth spline functions. As discussed in Section 3, we construct a prior on the spline coefficients, so that, for any given T variable, the linear and nonlinear effects corresponding to all associated M variables can be identified. Under these assumptions, the estimated coefficients of the T variables would be zero if both the linear and nonlinear effects of all associated M variables are zero. However, for a larger number of M variables, it is unlikely that both the linear and nonlinear effects of all M variables corresponding to any T variable are zero. If the number of Level 1 covariates (T variables) is also large, it becomes crucial to perform variable selection on them to enforce sparsity and interpretability.

In order to incorporate sparsity on the Level 1 covariates, one naïve approach could be to use discrete mixture priors such as spike-and-slab (George & McCulloch, 1993). But, since we consider the selection of the T covariates to be patient-specific, to apply spike-and-slab prior, we would need to assign a latent indicator for *each* coefficient for *every* patient. This approach would therefore substantially increase the number of parameters in the model. In addition, as discussed below, the spike-and-slab prior is not well-suited for functional regression coefficients. Instead, we rely on a Bayesian hard-thresholding approach where we truncate the coefficients with smaller absolute values to zeros so that only the important Level 1 variables are selected. For truncation, we take the Bayesian hard-thresholding function $h(z, t) = zI(|z| > t)$ to threshold the slope coefficients of the T variables. We modify the values of the slope coefficients given in Equation (3) as

$$\beta_j(\tau, \mathbf{M}_{ij}) = h\left(\sum_{k=1}^{q_j} f_{jk}^{(\tau)}(M_{ijk}), \lambda_j\right),$$

where λ_j is a "minimum effect size" for the effect of T_j to be considered as nonzero. Our use of the hard-thresholding

prior instead of the more commonly adopted Bayesian spike-and-slab prior is due to the functional nature of the regression coefficients $\beta_j(\cdot)$. The hard-thresholding prior allows selection at any given input whereas spike-and-slab is not able to handle an infinite-dimensional object. Another advantage of using the Bayesian hard-thresholding is that this "minimum effect size" can be set at any reasonable value by the user based on intuition or prior experience, or estimated by assigning a prior (as we discuss in the next section). Note that no hard-thresholding is performed (or needed) on the intercept term as our main interest is to perform variable selection among Level 1 covariates only.

2.3 | Likelihood construction

We now describe the likelihood construction the QUANTICO model, assimilating the above constructs, for any fixed quantile level of interest τ. We shorten the notation $\beta_0(\tau, \mathbf{M}_{i0})$ and $\beta_j(\tau, \mathbf{M}_{ij})$ in Equation (2) to $\beta_0^{(i)}(\tau)$ and $\beta_j^{(i)}(\tau)$, respectively, for $i = 1, \dots, n$, and $j = 1, \dots, p$. Following the principle of linear quantile regression (Koenkar & Bassett, 1978), $\beta_0^{(i)}(\tau)$ and $\beta_j^{(i)}(\tau)$ can be estimated by minimizing the loss function

$$V(\tau) = \sum_{i=1}^{n} \psi_\tau\left(y_i - \sum_{j=0}^{p} \beta_j^{(i)}(\tau) T_{ij}\right),$$

where $T_{i0} = 1$ and $\psi_\tau(t)$ is the check function given by $\psi_\tau(t) = \tau t$, if $t \geq 0$, or $\psi_\tau(t) = -(1-\tau)t$, if $t < 0$. Now consider the model

$$y_i = \sum_{j=0}^{p} \beta_j^{(i)}(\tau) T_{ij} + u_i, i = 1, \dots, n,$$

where u_i follows the i.i.d. asymmetric Laplace distribution with density $f(u|\tau) = \tau(1-\tau)\exp[-\psi_\tau(u)]$. Hence, the joint density of $y_1, \dots, y_n$ is given by

$$f(y_1, \dots, y_n|\tau) = \tau^n (1-\tau)^n \exp\left[-\sum_{i=1}^{n} \psi_\tau\left(y_i - \sum_{j=0}^{p} \beta_j^{(i)}(\tau) T_{ij}\right)\right].$$

Following Yu and Moyeed (2001), for any given τ, minimizing $V(\tau)$ with respect to $\beta_j^{(i)}(\tau)$'s for $j = 1, \dots, p$ is equivalent to maximizing $f(u_1, \dots, u_n|\tau)$. Hence, to estimate the τth quantile, the likelihood is given by

$$L_\tau(\{\beta_0^{(i)}, \dots, \beta_p^{(i)}\}_{i=1}^n | Y, T, M) = \tau^n (1-\tau)^n \exp\left[-\sum_{i=1}^{n} \psi_\tau\left(y_i - \sum_{j=0}^{p} \beta_j^{(i)}(\tau) T_{ij}\right)\right].$$

3 | PRIOR FORMULATIONS

Note that the intercept and the slope terms ($\beta_0^{(i)}(\tau)$ and $\beta_j^{(i)}(\tau)$, respectively) are functions of the spline coefficients $\boldsymbol{\alpha}_{jk}^{\tau}$. In this section, we describe the prior on the spline coefficients $\boldsymbol{\alpha}_{jk}^{\tau}$, the prior used to induce selection of Level 2 covariates, and the prior assigned to the thresholding parameter λ_j to select Level 1 covariates.

Prior on spline coefficients

We propose the use of a penalized spline in order to have a flexible but smooth fit. Specifically, we choose a large number of knots placed at equally spaced quantiles of the covariates so that the local features can be captured, and penalize the roughness of $f_{jk}^{\tau}(\cdot)$ through an improper Gaussian random walk prior on $\boldsymbol{\alpha}_{jk}^{\tau}$ given by $\boldsymbol{\alpha}_{jk}^{\tau} \sim N(\mathbf{0}, s\mathbf{K}^-)$. The penalty matrix $\mathbf{K}$ is constructed from the second-order differences of adjacent spline coefficients, which essentially penalizes the second derivatives of $f_{jk}^{\tau}(\cdot)$. Larger value of s leads to a smoother fit, while smaller value of s leads to an irregular fit (Ni et al., 2015). Note that Equation (3) is not identifiable since adding some quantity to any term of the summation and deducting the same quantity from any other term would yield the same summation value. In addition, $\mathbf{K}$ is singular and therefore no penalty is imposed on the linear and constant trend of $f_{jk}^{\tau}(\cdot)$ (i.e., the null space of $\boldsymbol{K}$). In order to alleviate these two issues, we consider a similar approach as in Scheipl et al. (2012) and transform the spline bases into orthonormal bases. Let $\mathbf{S}_{jk} = (\mathbf{S}_{1jk}, \dots, \mathbf{S}_{njk})$. For a given quantile τ, consider the spectral decomposition of the covariance of $\mathbf{S}_{jk}\boldsymbol{\alpha}_{jk}^{(\tau)}$

$$\text{cov}(\mathbf{S}_{jk}\boldsymbol{\alpha}_{jk}^{(\tau)}) = s\mathbf{S}_{jk}\mathbf{K}^-\mathbf{S}_{jk}^T = \begin{bmatrix}\mathbf{U}_{jk} & *\end{bmatrix}\begin{bmatrix}\mathbf{D}_{jk} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_{jk} & *\end{bmatrix}^T, \quad (4)$$

where $\mathbf{U}_{jk}$ is the orthonormal matrix of the eigenvectors corresponding to the positive eigenvalues in the diagonal matrix $\mathbf{D}_{jk}$. Let $\mathbf{S}_{jk}^* = \mathbf{U}_{jk}\mathbf{D}_{jk}^{\frac{1}{2}}$. Now if we assume an independent proper prior $\boldsymbol{\alpha}_{jk}^{*\tau} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, then the nonlinear part of $f_{jk}(\cdot)$ can be parameterized by $\mathbf{S}_{jk}^*\boldsymbol{\alpha}_{jk}^{*\tau}$ which has a proper distribution proportional to the distribution of the original improper prior $\mathbf{S}_{jk}\boldsymbol{\alpha}_{jk}^{\tau}$. Suppose we denote the effect size of the jth covariate for the ith subject before applying the hard-thresholding by $\beta_j^*(\tau, \mathbf{M}_{ij})$. Thus the full reparameterization of $\beta_j^*(\tau, \mathbf{M}_{ij})$ is now given by

$$\beta_j^*(\tau, \mathbf{M}_{ij}) = \sum_{k=1}^{q_j} f_{jk}^{(\tau)}(M_{ijk}) = \sum_{k=1}^{q_j} \mathbf{S}_{jk}^*\boldsymbol{\alpha}_{jk}^{*\tau} + \sum_{k=1}^{q_j} M_{ijk}\alpha_{jk}^{0\tau} + \alpha_j^{\tau}, \quad (5)$$

where α_j^{τ} is the global constant term absorbing all constant terms from splines. This parameterization adds the flexibility to separately shrink and estimate the linear, nonlinear, and constant effects of the Level 2 (M) variables on the coefficients of the Level 1 (T) variables. In order to make the proposed method computationally more efficient, we only consider the first several eigenvectors which explain at least 99.5% of the variation, a similar idea as in principal component analysis. In order to induce sparsity on the linear, nonlinear, and the constant effects, we consider a parameter-expanded normal mixture of inverse gamma (peNMIG) prior on each $\boldsymbol{\alpha}_{jk}^{*\tau}, \alpha_{jk}^{0\tau}$, and α_j^{τ} separately. We opt for the peNMIG prior since it is known to provide a more efficient Markov chain Monte Carlo (MCMC) algorithm compared to traditional spike-and-slab priors given the multivariate nature of the spline coefficients (Scheipl et al., 2012). peNMIG multiplicatively expands $\boldsymbol{\alpha}_{jk}^{*\tau}$ as $\boldsymbol{\alpha}_{jk}^{*\tau} = \eta_{jk}\boldsymbol{\xi}_{jk}$, where η_{jk} is a scalar parameter and $\boldsymbol{\xi}_{jk}$ is a vector of the same length as $\boldsymbol{\alpha}_{jk}^{*\tau}$. Each of $\alpha_{jk}^{0\tau}$ and α_j^{τ} is also expanded in a similar fashion. A brief discussion on the choice of such a prior is provided in Section A of the Supporting Information.

Priors for selection of Level 2 covariates

Since the Level 2 coefficients are at population level, we can induce explicit selection using a spike-and-slab prior on η_{jk},

$$\eta_{jk} \sim \gamma_{jk} N(0, t_{jk}) + (1 - \gamma_{jk}) N(0, v_0 t_{jk}), \tag{6}$$

where $\gamma_{jk} \sim Ber(\rho)$ and v_0 is a fixed very small quantity close to zero. The selection of η_{jk} as a nonzero effect is indicated by the binary variable γ_{jk}, and thus γ_{jk} indicates the selection of $\boldsymbol{\alpha}_{jk}^{*\tau}$ vector due to the multiplicative construction. In terms of interpretation, the binary variable $\gamma_{jk} = 1$ indicates that M_{ijk} has nonzero *nonlinear effect* on T_{ij}. Similarly, in the expansion of $\alpha_{jk}^{0\tau}$ and α_j^{τ}, $\gamma_{jk} = 1$ indicates the presence of *linear and constant effects* of M_{ijk} on T_{ij}, respectively. Thus, based on the estimated values of γ_{jk} in the expansion of $\boldsymbol{\alpha}_{jk}^{*\tau}$ and $\alpha_{jk}^{0\tau}$, we can identify the presence of nonlinear and linear effects of Level 2 covariates, respectively. In essence, this construction allows the flexibility of considering only linear or only nonlinear or joint effects simultaneously. We choose conjugate hyperpriors for t_{jk} and ρ, $t_{jk} \sim IG(a_t, b_t)$ and $\rho \sim Beta(a_\rho, b_\rho)$. As the number of Level 2 covariates increases, this Beta–Bernoulli prior automatically corrects for multiplicity by making the posterior distribution of ρ concentrated at small values near 0 (Scott & Berger, 2010). We assign a mixture normal prior on each element of the vector $\boldsymbol{\xi}_{jk} = (\xi_{jk}^{(i)})$, $\xi_{jk}^{(i)} \sim \frac{1}{2}N(1,1) + \frac{1}{2}N(-1,1)$. The structure of this assumed prior discourages small effects. In a similar fashion, peNMIG priors are assumed for $\alpha_{jk}^{0\tau}$ and α_j^{τ} as well.

Prior on hard-thresholding for selection of Level 1 covariates

As mentioned before, to select the Level 1 covariates, we adopt a hard-thresholding approach where we truncate the nonzero coefficient of the jth Level 1 covariate with absolute value less than λ_j to 0. Since sparsity is induced while estimating the linear, nonlinear, and constant effects of the Level 2 covariates on the Level 1 covariates, it is possible that $\beta_j(\cdot) = 0$ even before hard-thresholding (when $\boldsymbol{\alpha}_{jk}^{*\tau}$, $\alpha_{jk}^{0\tau}$, and α_j^{τ} are zero for all $k = 1, \ldots, q_j$). In that case, λ_j can take any value without affecting the resulting estimates. To resolve this identifiability issue, we take $\lambda_j = \lambda$ for $j = 1, \ldots, p$. In the presence of at least one nonzero $\beta_j(\cdot)$, λ is now well-defined. We put a gamma prior on λ, $\lambda \sim Gamma(a_\lambda, b_\lambda)$. The values of the shape and scale parameters (a_λ, b_λ) can be taken so that the mean of the gamma distribution (i.e., $a_\lambda b_\lambda$) is equal to the desired cutoff (based on intuition or prior experience). A brief discussion on how to choose the parameters of the gamma prior is provided in Supporting Information (Section A). Note that a more conventional variable selection prior, the spike-and-slab, is often used for finite-dimensional parametric models. We choose to use the random hard-thresholding because it is more suitable for infinite-dimensional functional selection, as discussed in Section 2.2. A schematic illustration of the proposed model and its key parameters is given in Figure 1C.

4 | MCMC AND POSTERIOR INFERENCE

The posterior distributions of the model parameters are not analytically tractable. Therefore, an MCMC sampling algorithm is required to generate samples from the posterior distribution. We use a Gibbs sampling scheme for updating parameters using their full conditional distributions (τ_{jk}, γ_{jk}, ρ) and Metropolis sampling scheme for the parameters without closed-form conditional distributions ($\eta_{jk}, \boldsymbol{\xi}_{jk}, \lambda$). Details on the full conditional distributions are provided in Section B of the Supporting Information.

Inferential summaries

When using this sampling algorithm, the quantile level of interest is fixed at the desired level. In order to estimate multiple quantiles simultaneously, the algorithm can be run in parallel for faster computation. The selection

of Level 1 and Level 2 covariates can be performed in several ways based on the marginal posterior probability of inclusion. For the selection of Level 1 covariates, we consider the cutoff of the posterior probability of inclusion to be 0.5. A similar approach of thresholding the posterior inclusion probabilities can be taken for the selection of Level 2 covariates. In the presence of a higher number of Level 2 covariates, a false discovery rate (FDR) controlling approach can also be considered (Baladandayuthapani et al., 2010); see Section C of the Supporting Information.

The proposed model allows the identification of both linear and nonlinear effect components of $\mathbf{M}_j$ within the effect size of T_j for all subjects. From the posterior samples, the posterior probabilities of both linear and nonlinear effect of Level 2 covariates within the effect size of Level 1 covariate can be calculated explicitly. Due to the induction of sparsity on the effect-size components of Level 2 covariates, the effect size of each Level 2 covariate can be categorized as one of four possible cases: linear, nonlinear, both, or none.

Based on the posterior mean estimate of the effect sizes of the Level 1 covariates over a grid of quantiles, patient-specific posterior credible intervals over the quantiles can be obtained using QUANTICO. To calculate the posterior 95% credible interval of the quantile function for the ith patient, we calculate the posterior mean of $Q_y(\tau_l|\mathbf{T}_{i\cdot}, \mathbf{M}_{i\cdot})$ over the quantile grid $\tau_l = 0.1l$ for $l = 1, \ldots, 9$. Then, for quantile level τ_l, we calculate the 95th percentile of the values $|Q_y^{(z)}(\tau_l|\mathbf{T}_{i\cdot}, \mathbf{M}_{i\cdot}) - \widehat{Q}_y(\tau_l|\mathbf{T}_{i\cdot}, \mathbf{M}_{i\cdot})|$ from the posterior sample, where $Q_y^{(z)}(\tau_l|\mathbf{T}_{i\cdot}, \mathbf{M}_{i\cdot})$ denotes the value of $Q_y(\tau_l|\mathbf{T}_{i\cdot}, \mathbf{M}_{i\cdot})$ obtained in the zth posterior sample. Thus, we derive the width of the posterior 95% credible interval at all τ_l's. Furthermore, by taking a dense grid of quantiles over [0,1], patient-specific uniform posterior credible intervals can be calculated. A detailed description on the calculation of posterior pointwise and uniform credible intervals is provided in Supporting Information Section D.

5 | SIMULATION STUDIES

In this section, we evaluate the variable selection performance of QUANTICO across both Level 1 and Level 2 covariates and illustrate the subject-specific estimation at different quantile levels using simulation studies. The performance of QUANTICO is compared with varying coefficient quantile regression model (VCQRM) and variable selection in quantile regression using the lasso penalty (LASSO-QR) (Wu & Liu, 2009) as well as its Bayesian alternatives. All of these methods are variants of quantile-based VCMs and we provide explicit details of each method in the Section C of the Supporting Information.

5.1 | Variable selection performance

To compare the variable selection performance across both Level 1 (T) and Level 2 (M) covariates, we calculate the true positive rate (TPR), false positive rate (FPR), and area under receiver operating characteristic (ROC) curve (AUC) separately for the T and M variables. A detailed description of the simulation design and computation details of the metrics is provided in Supporting Information Section C.

Simulation structure

To compare the performance of QUANTICO, VCQRM, and LASSO-QR, we consider two scenarios with sample sizes of $n = 100$ and $n = 200$. In order to understand how the selection of the T variables is carried out in the presence of a higher number of covariates, we consider the cases $p = 5, 10$ for sample size scenario $n = 100$ and the cases $p = 5, 10, 20$ for the sample size scenario $n = 200$. For the cases where $p > 5$, the true model remains as given by Equation (1) of the Supporting Information. So, the additional T covariates considered in $p > 5$ scenarios have no true effect on Y. The simulation is repeated 25 times, and each time new data are generated from the quantile function given in Equation (1) of the Supporting Information. The mean and the standard deviation of the TPR, FPR, and AUC for the T and M variables are computed separately at $\tau = 0.1, 0.5, 0.9$ for all the methods. In addition, for a few high-dimensional scenarios the performance of QUANTICO is evaluated in Section F of the Supporting Information. We observe that QUANTICO maintains a high TPR and AUC for larger values of p and q_j as well, along with low FPR rates. We also evaluate the performance of QUANTICO compared to two other existing quantile regression methods, implemented in the R packages `rqPen` (Sherwood & Maidman, 2019) and `Brq` (Alhamzawi & Ali, 2020) which is also provided in Section F of the Supporting Information. It is observed that QUANTICO outperforms `rqPen` and `Brq`, in general.

Comparative performance evaluation

The comparative performance of the three methods is reported in Table 1. In general, QUANTICO and VCQRM perform better than LASSO-QR. QUANTICO results in better selection of the T (Level 1) variables, and a large improvement in terms of FPR and AUC over VCQRM. In VCQRM, we do not incorporate any thresholding of the slope terms. Although it is possible to have a zero slope or intercept term without thresholding if all estimated linear,

TABLE 1 The result for the method(s) with the best performance is marked in bold, comparative performance study of QUANTICO, varying coefficient quantile regression model (VCQRM) and LASSO quantile regression (LASSO-QR) methods at quantile levels $\tau = 0.1, 0.5, 0.9$ for different sample size and number of T variable scenarios. lvl1TPR, lvl1FPR, and lvl1AUC denote the true positive rate, false positive rate, and area under receiver operating characteristic curve (AUC) of Level 1 covariates; and lvl2TPR, lvl2FPR, and lvl2AUC denote the same for Level 2 covariates

(n, p)	Measures	$\tau = 0.1$			$\tau = 0.5$			$\tau = 0.9$		
		QUANTICO	VCQRM	LASSO-QR	QUANTICO	VCQRM	LASSO-QR	QUANTICO	VCQRM	LASSO-QR
(100,5)	lvl1TPR	0.94 (0.22)	**1.00 (0.00)**	**1.00 (0.00)**	0.97 (0.10)	**1.00 (0.00)**	**1.00 (0.00)**	0.85 (0.20)	**1.00 (0.00)**	0.99 (0.07)
	lvl1FPR	**0.00 (0.01)**	1.00 (0.00)	0.56 (0.41)	**0.24 (0.18)**	1.00 (0.00)	0.65 (0.33)	**0.03 (0.11)**	1.00 (0.00)	0.66 (0.40)
	lvl1AUC	**0.97 (0.11)**	0.50 (0.00)	0.94 (0.16)	**0.95 (0.08)**	0.50 (0.00)	0.93 (0.16)	**0.94 (0.09)**	0.50 (0.00)	0.91 (0.18)
	lvl2TPR	**0.10 (0.25)**	0.08 (0.19)	NA	**0.58 (0.34)**	0.48 (0.37)	NA	**0.32 (0.38)**	0.24 (0.39)	NA
	lvl2FPR	**0.00 (0.00)**	**0.00 (0.00)**	NA	**0.05 (0.04)**	**0.05 (0.05)**	NA	**0.01 (0.01)**	**0.01 (0.01)**	NA
	lvl2AUC	**0.96 (0.09)**	0.91 (0.13)	NA	0.88 (0.15)	**0.89 (0.11)**	NA	0.85 (0.14)	**0.90 (0.13)**	NA
(100,10)	lvl1TPR	0.90 (0.25)	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.01)**	**1.00 (0.00)**	0.98 (0.10)	0.78 (0.25)	**1.00 (0.00)**	0.99 (0.07)
	lvl1FPR	**0.01 (0.03)**	1.00 (0.00)	0.42 (0.28)	**0.13 (0.09)**	1.00 (0.00)	0.62 (0.30)	**0.06 (0.10)**	1.00 (0.00)	0.59 (0.32)
	lvl1AUC	**0.98 (0.08)**	0.50 (0.00)	0.96 (0.09)	**0.99 (0.02)**	0.50 (0.00)	0.96 (0.10)	0.90 (0.12)	0.50 (0.00)	**0.91 (0.10)**
	lvl2TPR	**0.06 (0.17)**	0.04 (0.14)	NA	**0.42 (0.37)**	0.40 (0.35)	NA	0.12 (0.30)	**0.18 (0.32)**	NA
	lvl2FPR	**0.00 (0.00)**	**0.00 (0.00)**	NA	**0.02 (0.02)**	**0.02 (0.02)**	NA	**0.00 (0.01)**	**0.00 (0.00)**	NA
	lvl2AUC	0.90 (0.14)	**0.91 (0.09)**	NA	0.86 (0.15)	**0.88 (0.13)**	NA	0.81 (0.18)	**0.86 (0.16)**	NA
(200,5)	lvl1TPR	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.00)**	0.99 (0.03)	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.00)**
	lvl1FPR	**0.00 (0.00)**	1.00 (0.00)	0.57 (0.35)	**0.19 (0.16)**	1.00 (0.00)	0.73 (0.24)	**0.02 (0.09)**	1.00 (0.00)	0.78 (0.36)
	lvl1AUC	**1.00 (0.00)**	0.50 (0.00)	0.99 (0.03)	0.88 (0.20)	0.50 (0.00)	**0.98 (0.07)**	**1.00 (0.02)**	0.50 (0.00)	0.99 (0.04)
	lvl2TPR	**0.98 (0.10)**	**0.98(0.10)**	NA	0.86 (0.23)	**0.88 (0.22)**	NA	0.88 (0.30)	**0.96 (0.20)**	NA
	lvl2FPR	**0.03 (0.01)**	**0.03 (0.01)**	NA	**0.10 (0.06)**	**0.10 (0.05)**	NA	**0.03 (0.02)**	**0.03 (0.02)**	NA
	lvl2AUC	**1.00 (0.00)**	**1.00 (0.00)**	NA	**0.95 (0.01)**	0.94 (0.11)	NA	**1.00 (0.01)**	**1.00 (0.01)**	NA
(200,10)	lvl1TPR	**1.00 (0.00)**	**1.00 (0.00)**	**1.00 (0.00)**	0.99 (0.03)	**1.00 (0.00)**	**1.00 (0.00)**	0.99 (0.04)	**1.00 (0.00)**	**1.00 (0.00)**
	lvl1FPR	**0.00 (0.00)**	1.00 (0.00)	0.44 (0.28)	**0.16 (0.09)**	1.00 (0.00)	0.55 (0.31)	**0.01 (0.02)**	1.00 (0.00)	0.59 (0.26)
	lvl1AUC	**1.00 (0.00)**	0.50 (0.00)	0.99 (0.03)	**0.98 (0.03)**	0.50 (0.00)	0.97 (0.10)	**1.00 (0.00)**	0.50 (0.00)	0.98 (0.03)
	lvl2TPR	**1.00 (0.00)**	0.94 (0.22)	NA	**0.82 (0.24)**	0.80 (0.25)	NA	**0.92 (0.19)**	0.88 (0.26)	NA
	lvl2FPR	0.02 (0.00)	**0.01 (0.01)**	NA	0.06 (0.03)	**0.05 (0.03)**	NA	**0.01 (0.01)**	**0.01 (0.01)**	NA
	lvl2AUC	**1.00 (0.00)**	**1.00 (0.00)**	NA	**0.95 (0.08)**	**0.95 (0.09)**	NA	**1.00 (0.01)**	0.99 (0.04)	NA
(200,20)	lvl1TPR	**1.00 (0.01)**	**1.00 (0.00)**	**1.00 (0.00)**	0.93 (0.22)	**1.00 (0.00)**	**1.00 (0.00)**	0.97 (0.13)	**1.00 (0.00)**	**1.00 (0.00)**
	lvl1FPR	**0.00 (0.00)**	1.00 (0.00)	0.28 (0.22)	**0.11 (0.06)**	1.00 (0.00)	0.34 (0.25)	**0.01 (0.02)**	1.00 (0.00)	0.43 (0.22)
	lvl1AUC	**0.99 (0.01)**	0.50 (0.00)	**0.99 (0.02)**	0.92 (0.22)	0.50 (0.00)	**1.00 (0.01)**	**0.99 (0.06)**	0.50 (0.00)	0.98 (0.03)
	lvl2TPR	**0.96 (0.20)**	0.92 (0.24)	NA	**0.70 (0.35)**	0.68 (0.28)	NA	**0.78 (0.33)**	**0.78 (0.33)**	NA
	lvl2FPR	**0.01 (0.00)**	**0.01 (0.00)**	NA	0.03 (0.02)	**0.02 (0.01)**	NA	0.01 (0.01)	**0.00 (0.00)**	NA
	lvl2AUC	**1.00 (0.00)**	**1.00 (0.00)**	NA	0.88 (0.18)	**0.95 (0.08)**	NA	**0.98 (0.06)**	**0.98 (0.09)**	NA

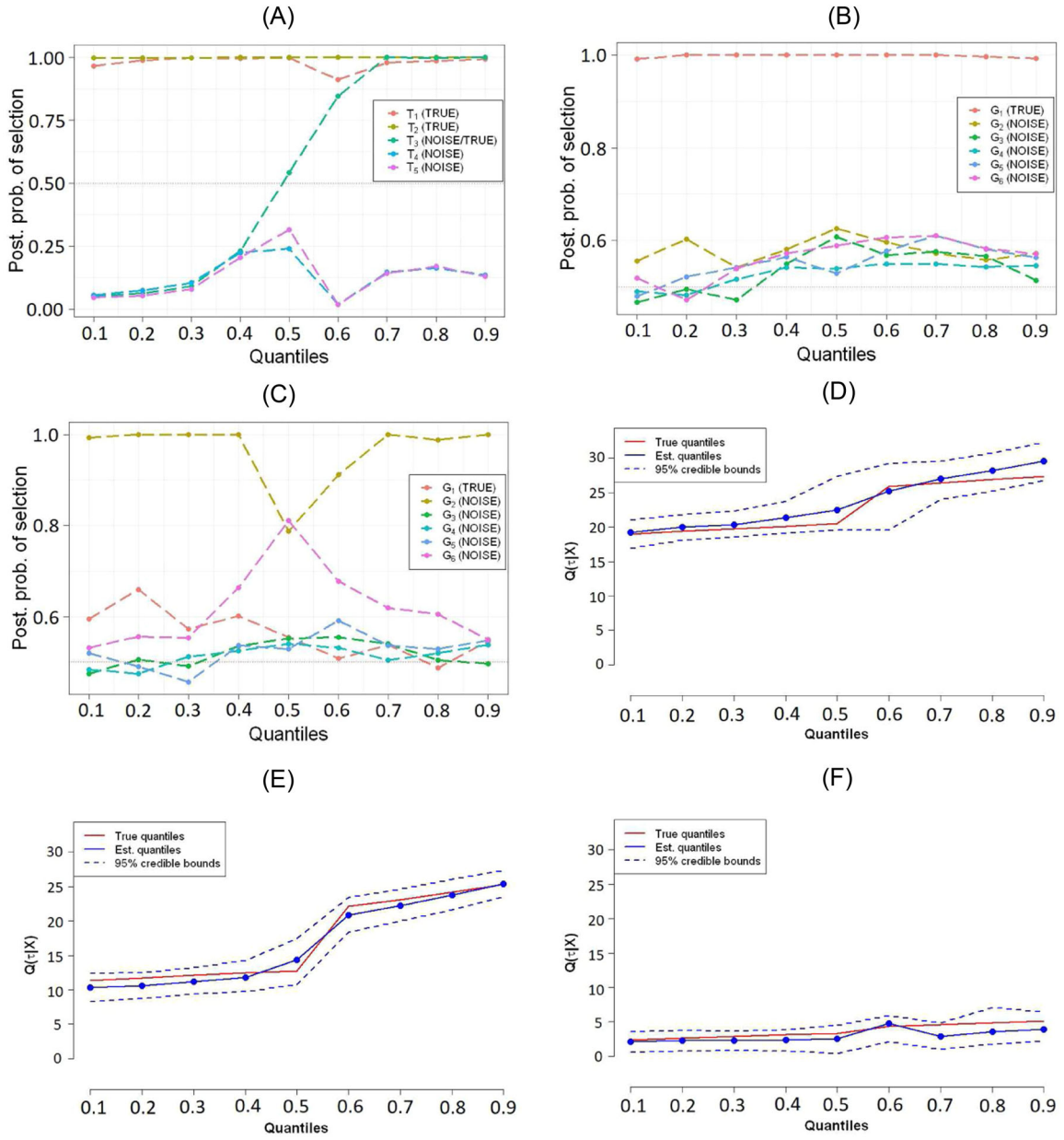

FIGURE 2 Plots from the simulation study for ($n = 200$, $p = 6$) scenario using QUANTICO: (A) Posterior probability of selection of T variables, where the true quantile function depends on T_1, T_2 for all quantile levels $\tau \in [0, 1]$ and on T_3 for quantiles $\tau > 0.5$. T_4, T_5 are noise variables. (B) Probability of selection of second-level covariates for T_1, the true coefficient of P_1 only involves G_1. (C) Probability of selection of second-level covariates for T_2, The true coefficient of P_2 only involves G_2. (D–F) True and estimated quantile functions corresponding to three randomly selected simulated subjects, 95% credible bounds are also shown

nonlinear, and constant effects of all Level 2 covariates are zero, in the simulation results, the FPR of Level 1 covariates came out to be 1. We do not observe any strong pattern of differences in the comparative performance of QUANTICO and VCQRM in terms of TPR, FPR, and AUC for M (Level 2) variables, since they follow same mechanism for selection of M variables. In terms of performance, in Table 1, we report that QUANTICO yields very low TPR as well as FPR for Level 2 covariates for sample size scenario $n = 100$ despite a high AUC, which implies that QUANTICO is very conservative in its estimation. In QUANTICO, the selection of Level 2 covariates is performed using an FDR-based approach with cutoff $\alpha = 0.2$. The performance of QUANTICO improves as n increases (as noted in Table 1).

We further report the selection and estimation performance of QUANTICO evaluated at quantile levels $\tau = 0.1, 0.2, \ldots, 0.9$. Figure 2A illustrates the selection of

T variables. QUANTICO selected the correct T variables across all quantiles. Specifically, it selected variable T_3 for the quantiles greater than 0.5 as it is in the true model. In Figure 2B,C, the corresponding true G variables have the highest posterior probability of selection for T_1 and T_2, respectively. In Figure 2D–F, the true and estimated quantile functions are plotted for three randomly selected subjects along with the corresponding estimated 95% credible bounds.

We also compute the uniform posterior 95% credible intervals of the quantile functions. To improve coverage of the uniform credible intervals, we increase the width using an inflation factor. Although for point estimates, the observed coverage might come close to the percentage of the computed credible interval, for functions, in practice, it is a common phenomenon that the observed coverage may be less than the actual percentage of the credible interval. This undercoverage of the uniform posterior credible band of smooth functions is a well-known property and has been addressed in several articles (Cox, 1993; Das & Ghosal, 2017; Knapik et al., 2011; Szabo et al., 2015). In order to improve coverage, either undersmoothing or inflation of the obtained credible interval is required (Yoo & Ghosal, 2016). To improve coverage of the posterior uniform credible intervals, we increase the width using an inflation factor. As mentioned in Remark 5.4 and Theorem 5.3 in Yoo and Ghosal (2016), the inflation factor of the uniform credible interval should be taken such that it slowly increases to infinity as a function of sample size. Through experimentation, we observe that the inflation factor $f(n) = 1.5\sqrt{\log(n)}$ (which has a similar form to that considered in Das and Ghosal (2017) in the context of uniform credible bound over quantiles) works well in simulation under various settings for the sample size and number of Level 1 covariates. A detailed discussion regarding the computation of the uniform posterior 95% credible intervals of the quantile functions along with an extensive study on the computational time is provided in Supporting Information Section D.

6 | CASE STUDY ON IMMUNE ARCHITECTURE OF NON-SMALL-CELL LUNG CANCER (NSCLC)

6.1 | Scientific problem and data description

Lung cancer is the second most common cancer and is one of the leading causes of cancer death. Though early stage patients can be treated with surgery, late stage lung cancer requires the use of systemic therapies (Ettinger et al., 2017). In recent years, immunotherapies have emerged as a successful treatment in a subset of late stage lung cancers, largely through their ability to boost the activity of T-cells, the subset of immune cells that target infected or malignant cells based on the detection of specific antigens. However, a large proportion of lung cancer patients still do not respond to immunotherapy (Doroshow et al., 2019).

The nature of the antigens specifically recognized by T-cells has been the object of intense focus, and recent work has suggested that somatic mutations harbored by the tumor can be presented to T-cells as neoantigens (Lee et al., 2018). Analysis of specific somatic mutations and overall tumor mutational burden (TMB) has shown a positive association with response to immunotherapy in patients with NSCLC. This supports the role of these mutations in aiding T-cell responses by increasing tumor immunogenicity. Recent technologies have emerged to sequence the TCR to gain insight into T-cell responses, and studies have confirmed that TCR sequencing can be used to monitor immune response in various types of cancer (Page et al., 2016). In order to develop patient-specific immunotherapeutic treatment strategies for NSCLC, it is of critical importance to understand the interplay between somatic mutations, TCR variables, and the immune microenvironment, illustrated schematically in Figure 1A.

We consider a cohort of 215 NSCLC patients recruited at UT MD Anderson Cancer Center. The tumors of these subjects were analyzed to obtain immune profiling, TCR sequencing, and mutational status of immune-related genes. We focus here on understanding the impact of mutation and TCR variables on the CD8 marker, which is the outcome variable in our analysis and is discussed in more detail below. In order to assess the effect of the explanatory variables on the outcome variable across different patients, we would like to estimate the patient-specific effect at different quantiles of the response.

As our Level 1 covariates, we take standard summary measures of TCR sequencing data, including T-cell clonality, T-cell entropy, T-cell productive proportion of cells, and T-cell richness. T-cell clonality is a measure of heterogeneity among T-cells and has been linked to patient outcomes (Reuben et al., 2017). T-cell entropy is Shannon's entropy, and highly diverse samples have comparatively higher entropy. The T-cell productive proportion of cells denotes the proportion of the tumor that consists of T-cells. TCR richness is another measure of TCR heterogeneity, defined as the number of different unique sequences in the sample.

As our Level 2 covariates, we consider mutation counts for the top six most frequently mutated genes across the cohort, namely, CSMD3, MUC16, RYR2, TP53, TTN, and USH2A, along with total TMB, which is the total number of mutations observed per sample. As the mutation counts for individual genes have sparse values (i.e., zeros for a large

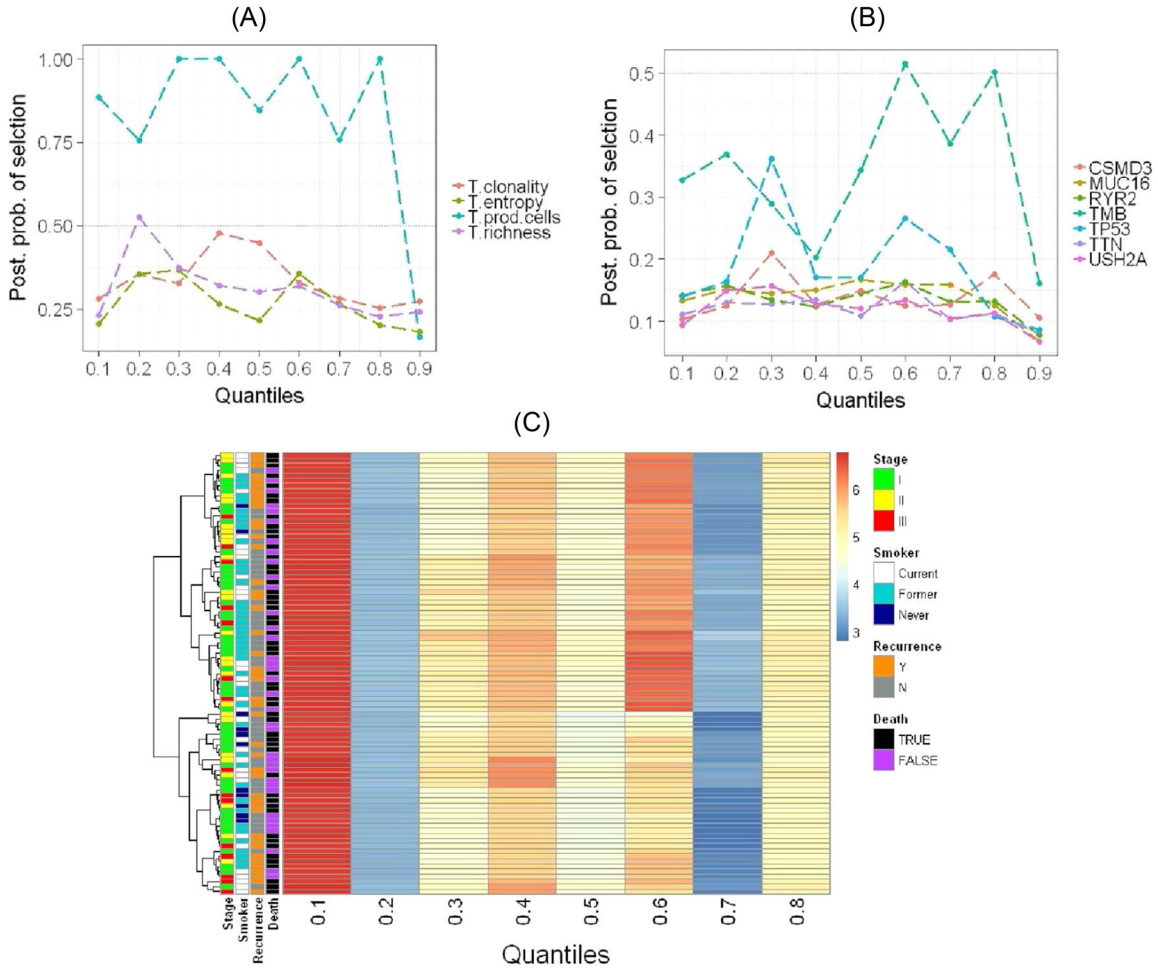

FIGURE 3 (A) Posterior probability of selection of TCR variables based on average of individual posterior probability of the same for each of the 87 patients. T-cell productive proportion is the only selected variable for quantiles $\tau = 0.1, 0.2, \ldots, 0.8$. No TCR variables are selected at $\tau = 0.9$. (B) Posterior probability of mutation variables having nonzero effect on the coefficient of the T-cell productive proportion variable. Tumor Mutation Burden (TMB) is shown to have highest nonzero effect among all mutation variables, specially at higher quantile effects. (C) Hierarchical clustering of the patient-specific estimated coefficients of T-cell productive proportion of cells at quantile levels $\tau = 0.1, 0.2, \ldots, 0.8$ for 87 considered NSCLC patients. Three clinical variables, that is, smoking status, recurrence, and vital status are also shown

proportion of patients), we only consider the linear and constant effects for those six variables. For TMB, linear, nonlinear, and constant effects are allowed.

As our response variable, we focus on CD8 abundance as a key measure of immune activity. CD8 is a protein found on the surface of cytotoxic T-cells. CD8+ T-cells have the ability to mount a response against pathogens and defend against tumors by killing transformed tumor cells (Berg & Forman, 2006), and are therefore a vital part of cancer immunity. In precision oncology, understanding the patient-specific effect of T-cell architecture on CD8+ T-cell abundance is therefore crucial.

Out of the cohort of 215 NSCLC patients, we focus on the 87 patients for which all three data types (immune, TCR, and mutational profiling) are available. Since the set of TCR variables considered have differing magnitudes, it is crucial to transform them to a similar range of values to make any comparison of their effect sizes meaningful. The same applies for the mutation variables. Therefore, before applying QUANTICO, we transform the TCR variables and the CD8 immune markers to unit intervals using log-normal cumulative distribution function (cdf) transformation (see Supporting Information Section G). The mutation variables are transformed into the unit interval using a linear transformation.

6.2 | Results

Population level findings

Using the proposed method, the coefficients of the TCR variables are estimated for each patient at nine quantile levels $\tau = 0.1, 0.2, \ldots, 0.9$. We use the same values of the

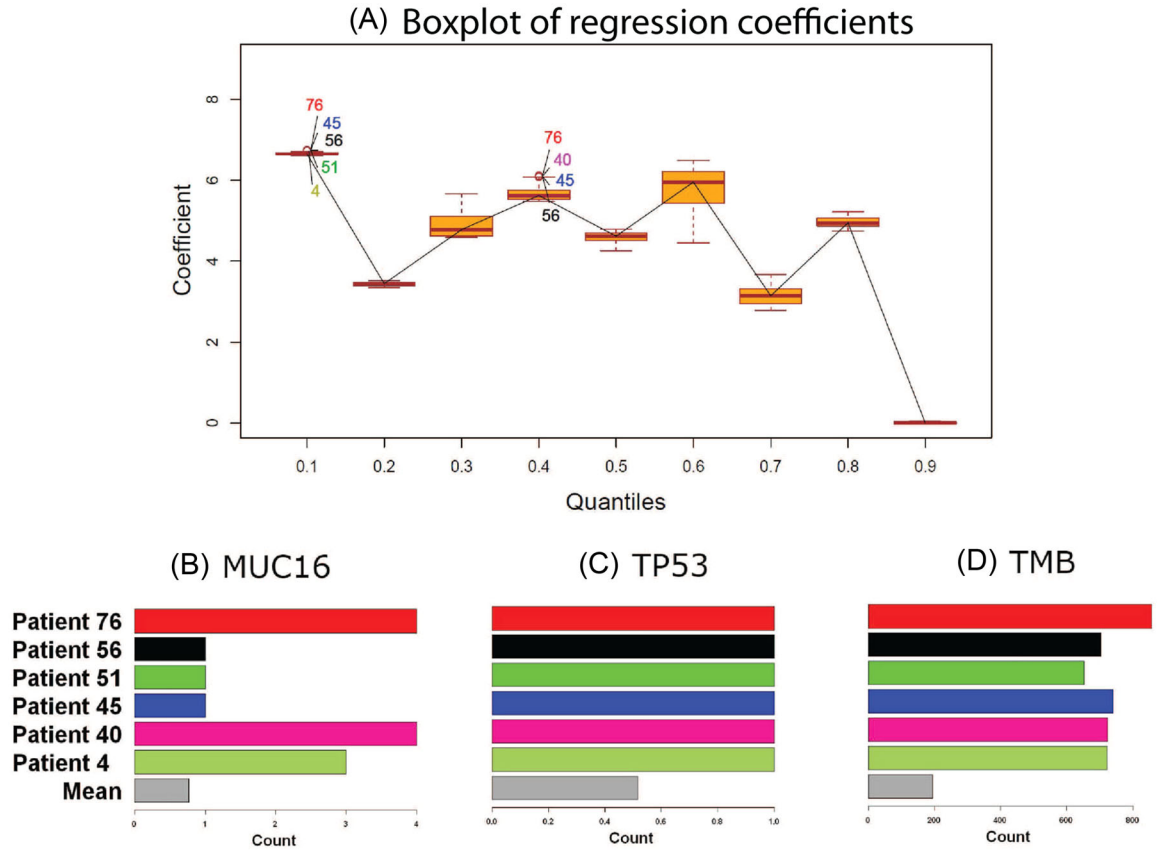

FIGURE 4 (A) Boxplot of the coefficient of productive proportion of T-cell for all the patients at quantile levels $\tau = (0.1, 0.2, \ldots, 0.9)$. The outlier patients are identified from the boxplot. (B–D) Barplot of the mutation counts of the genes MUC16, TP53, and tumor mutation burden (TMB) of the outlier patients and its mean value for all patients

hyperparameters as in the simulation study except the total number of iterations and burn-in, which are taken to be 50,000 and 10,000, respectively. The average posterior probability of selection of all TCR variables is plotted in Figure 3A. The T-cell productive proportion variable has an average posterior probability greater than 0.5 at all quantile levels except at $\tau = 0.9$, and T-cell entropy has an average posterior probability marginally higher than 0.5 at $\tau = 0.2$. In general, the average posterior probability of the TCR variables having a nonzero effect on the CD8+ immune cell abundance decreases at higher quantile levels. This implies that for patients with a lower abundance of CD8+ cells, the number of CD8+ cells has a stronger dependence on the TCR variable measures. However, in patients with a higher density of CD8+ cells, this dependence is less prominent.

To summarize our Level 2 findings, we assessed the effect of the mutation variables on the coefficient of the T-cell productive proportion, which was identified to be the most important among the TCR variables (Figure 3B). The nonzero effects of the mutation variables in the coefficient of the T-cell productive proportion are not strong for lower quantile levels, while TMB is shown to have a marginally higher posterior probability of having a nonzero effect at higher quantiles. Thus, at higher quantiles, a large proportion of the effect of the T-cell productive proportion on CD8+ immune cells is due to TMB. This is consistent with previous studies that have shown a positive correlation between TMB and CD8+ in melanoma (Reuben et al., 2017), where immunotherapy using checkpoint inhibitors have shown to be successful. Hence, our findings suggest that TMB could be used for predicting the response to anti-PD-1/PD-L1 therapies in NSCLC.

Patient-level findings

Our results also provide insights into patient-specific immune profiles offering the potential to guide the development of precision immuno-oncology treatment strategies based on patient-specific information such as patient smoking history and mutational profiles. As an illustration of this, in Figure 3C, we show the estimated coefficients of the T-cell productive proportions for each subject across quantiles in the rows of a heatmap, along with patient-level covariates. In this figure, we rely on hierarchical

clustering over the estimated coefficients at quantile levels $\tau = 0.1, \ldots, 0.8$ to group patients with similar coefficients. Focusing on the last few rows of the heatmap, it is apparent that patients with overall lower values of the coefficient of the T-cell productive proportion have higher recurrence rates of NSCLC. Layering this analysis with clinical covariates reveals that the effect size of T-cell productive proportion is in general lower for nonsmokers compared to recent and former smokers (Figure S3). Reuben et al. (2020) found higher T-cell clonality in current and former smokers compared to never smokers.

In Figure 4A, we show a boxplot of the coefficients of the T-cell productive proportion for all patients, and we detect outlier patients (patient numbers 4, 45, 51, 56, and 76) at quantile level 0.1 and an overlapping set of outliers (patient numbers 40, 45, 56, and 76) at quantile level 0.3. Since this coefficient is modeled as a function of the mutation variables, we further compare the number of mutations for these six distinct patients (Figures 4B–D). In general, these outlier patients have a higher number of mutations compared to the average value in the patient cohort; specifically for TMB, TP53, and MUC16 genes.

7 | CONCLUSIONS AND FUTURE WORK

In this paper, we propose QUANTICO with multilevel covariates where, at any specific quantile level, selection over the direct (Level 1) covariates is performed for each subject. A novel feature of the proposed model is the development of a quantile-specific varying sparsity coefficient estimation approach which allows us to explicitly delineate how at different quantiles, the response variable depends on different covariates for each subject. The proposed method also enables selection of the indirect (Level 2) covariates, and can be used for obtaining patient-specific posterior credible bands over the quantile levels.

The proposed method is used to analyze how the CD8 immune marker depends on TCR and mutation variables at different quantiles. We find that T-cell productive proportion is the most important TCR variable, and influences CD8 immune cells for most quantile levels. Out of all mutation variables considered, total TMB is found to be most important. Based on the structure of the relationship between the TCR and immune variable, we identify outlier patients, who turn out to have a higher number of mutations across several critical genes of known clinical relevance in cancer. This information is potentially useful to devise effective immunologic therapies for such patients(s) based on their unique immune architectures.

There are several potential refinements that could be made to our modeling framework. We can extend our approach to simultaneous quantile regression, where instead of estimating the model parameters at specific quantiles, the entire quantile function and associated parameters are estimated simultaneously (Das & Ghosal, 2018; Yang & Tokdar, 2017). In terms of theoretical excursions, one could investigate for such hierarchical VCQRMs, building on some of the theoretical results proposed for Bayesian quantile regression by ourselves and others (Das & Ghosal, 2017; Yang & Tokdar, 2017) regarding posterior consistency, rates of convergence and posterior contraction. Given the nontrivial nature of these explorations, we leave them as future work.

ACKNOWLEDGMENTS

VB was supported by NIH grants R01-CA160736, R01CA244845-01A1, R21-CA220299, and P30 CA46592, NSF Grant 1463233, and start-up funds from the U-M Rogel Cancer Center and School of Public Health. CBP was supported by NIH/NCI CCSG P30CA016672 and CPRIT Grant RP150521. KAD was partially supported by a CCSG NCI Grant P30 CA016672, NIH Grants UL1TR003167, 5R01GM122775, the prostate cancer SPORE P50 CA140388, CPRIT Grant RP160693, and the Moon Shots funding at MD Anderson Cancer Center. YN was partially supported by the NSF DMS-1918851 and NSF DMS-2112943.

OPEN RESEARCH BADGES

This article has earned an Open Materials badge for making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis. All materials are available at http://re3data.org/.

DATA AVAILABILITY STATEMENT

Code and data: Code for the simulations and real data analysis, along with a synthetic data set designed to closely resemble our example T-cell receptor (TCR) data, are available online at https://github.com/bayesrx/QUANTICO.

ORCID

Priyam Das https://orcid.org/0000-0003-2384-0486
Christine B. Peterson https://orcid.org/0000-0003-3316-0468
Veerabhadran Baladandayuthapani https://orcid.org/0000-0001-9107-3157

REFERENCES

Alhamzawi, R. & Ali, H. (2020) Brq: an R package for Bayesian quantile regression. *METRON*, 78(3), 313–328.

Baladandayuthapani, V., Ji, Y., Talluri, R., Nieto-Barajas, L.E. & Morris, J.S. (2010) Bayesian random segmentation models to

identify shared copy number aberrations for array CGH data. *Journal of the American Statistical Association*, 105(492), 1358–1375.

Berg, R.E. & Forman, J. (2006) The role of CD8 T cells in innate immunity and in antigen non-specific protection. *Current Opinion in Immunology*, 18(3), 338–343.

Biller, C. & Fahrmeir, L. (2001) Bayesian varying-coefficient models using adaptive regression splines. *Statistical Modelling*, 1(3), 195–211.

Cox, D.D. (1993) An analysis of Bayesian inference for non-parametric regression. *Annals of Statistics*, 21(2), 903–923.

Das, P. & Ghosal, S. (2017) Bayesian quantile regression using random B-spline series prior. *Computational Statistics & Data Analysis*, 109, 121–143.

Das, P. & Ghosal, S. (2018) Bayesian non-parametric simultaneous quantile regression for complete and grid data. *Computational Statistics & Data Analysis*, 127(C), 172–186.

Doroshow, D.B., Sanmamed, M.F., Hastings, K., Politi, K., Rimm, D.L., Chen, L. et al. (2019) Immunotherapy in non-small cell lung cancer: facts and hopes. *Clinical Cancer Research*, 25(15), 4592–4602.

Ettinger, D.S., Wood, D.E., Aisner, D.L., Akerley, W., Bauman, J., Chirieac, L.R. et al. (2017) Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, 15, 504–535.

Fan, J. & Zhang, W. (1999) Statistical estimation in varying coefficient models. *Annals of Statistics*, 27(5), 1491–1518.

George, E.I. & McCulloch, R.E. (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889.

Hastie, T. & Tibshirani, R. (1993) Varying-coefficient models. *Journal of the Royal Statistical Society: Series B*, 55, 757–796.

Honda, T. (2004) Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, 121, 113–125.

Kakimi, K., Karasaki, T., Matsushita, H. & Sugie, T. (2017) Advances in personalized cancer immunotherapy. *Breast Cancer*, 24(1), 16–24.

Kim, M. (2007) Quantile regression with varying coefficients. *Annals of Statistics*, 35(1), 92–108.

Knapik, B., van der Vaart, A. & van Zanten, J. (2011) Bayesian inverse problems with Gaussian priors. *Annals of Statistics*, 39(5), 2626–2657.

Koenkar, R. & Bassett, G. (1978) Regression quantiles. *Econometrica*, 46, 33–50.

Koslovsky, M.D., Hoffman, K.L., Daniel, C.R. & Vannucci, M. (2020) A Bayesian model of microbiome data for simultaneous identification of covariate associations and prediction of phenotypic outcomes. *Annals of Applied Statistics*, 14(3), 1471–1492.

Kottas, A. & Gelfand, A.E. (2001) Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96, 1458–1468.

Lee, C., Yelensky, R., Jooss, K. & Chan, T.A. (2018) Update on tumor neoantigens and their utility: why it is good to be different. *Trends in Immunology*, 39(7), 536–548.

Ni, Y., Stingo, F.C. & Baladandayuthapani, V. (2015) Bayesian nonlinear model selection for gene regulatory networks. *Biometrics*, 71(3), 585–595.

Ni, Y., Stingo, F.C., Ha, M.J., Akbani, R. & Baladandayuthapani, V. (2018) Bayesian hierarchical varying-sparsity regression models with application to cancer proteogenomics. *Journal of the American Statistical Association*, 114(525), 48–60.

Page, D.B., Yuan, J., Redmond, D., Wen, Y.H., Durack, J.C., Emerson, R. et al. (2016) Deep sequencing of T-cell receptor DNA as a biomarker of clonally expanded TILs in breast cancer after immunotherapy. *Cancer Immunology Research*, 4, 835–844.

Park, B.U., Mammen, E., Lee, Y.K. & Lee, E.R. (2013) Varying coefficient regression models: a review and new developments. *International Statistical Review*, 83(1), 36–64.

Reich, B.J. (2012) Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4), 535–553.

Reuben, A., Gittelman, R., Gao, J., Zhang, J., Yusko, E.C., Wu, C. et al. (2017) TCR repertoire intratumor heterogeneity in localized lung adenocarcinomas: an association with predicted neoantigen heterogeneity and postsurgical recurrence. *Cancer Discovery*, 7(10), 1088–1097.

Reuben, A., Spencer, C.N., Prieto, P.A., Gopalakrishnan, V., Reddy, S.M., Miller, J.P. et al. (2017) Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. *NPJ Genomic Medicine*, 2(1), 1–11.

Reuben, A., Zhang, J., Chiou, S., Gittelman, R.M., Li. J., Lee, W. et al. (2020) Comprehensive T cell repertoire characterization of non-small cell lung cancer. *Nature Communications*, 11(603), 1–13.

Scheipl, F., Fahrmeir, L. & Kneib, T. (2012) Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association*, 107(500), 1518–1532.

Scott, J.G. & Berger, J.O. (2010) Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Annals of Statistics*, 38(5), 2587–2619.

Sherwood, B. & Maidman, A. (2019) *Penalized quantile regression*. Available at: https://cran.r-project.org/web/packages/rqPen/rqPen.pdf [Accessed 19 May 2022].

Stingo, F.C., Guindani, M., Vannucci, M. & Calhoun, V.D. (2013) An integrative Bayesian modeling approach to imaging genetics. *Journal of the American Statistical Association*, 108(503), 876–891.

Szabo, B., van der Vaart, A. & van Zanten, J. (2015) Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Annals of Statistics*, 43(4), 1391–1428.

Tang, Q. & Wang, J. (2005) l_1-Estimation for varying coefficient model. *Statistics*, 39, 389–404.

Waldman, A.D., Fritz, J.M. & Lenardo, M.J. (2020) A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*, 20, 651–668.

Walsh, R. & Soo, R. (2020) Resistance to immune checkpoint inhibitors in non-small cell lung cancer: biomarkers and therapeutic strategies. *Therapeutic Advances in Medical Oncology*, 12.

Wang, H. & Xia, Y. (2009) Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104, 747–757.

Wu, Y. & Liu, Y. (2009) Variable selection in quantile regression. *Statistica Sinica*, 19, 801–817.

Yang, Y. & Tokdar, S. (2017) Joint estimation of quantile planes over arbitrary predictor spaces. *Journal of the American Statistical Association*, 112(519), 1107–1120.

Yoo, W.W. & Ghosal, S. (2016) Supremum norm posterior contraction and credible sets for non-parametric multivariate regression. *Annals of Statistics*, 44(3), 1069–1102.

Yu, K. & Moyeed, R. (2001) Bayesian quantile regression. *Statisics and Probabilty Letters*, 54, 437–447.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2– 6, along with relevant codes are available with this paper at the *Biometrics* website on Wiley Online Library. The codes are also available online at https://github.com/bayesrx/QUANTICO.

Supporting Information

How to cite this article: Das, P., Peterson, C.B., Ni, Y., Reuben, A., Zhang, J., & Zhang, J. et al. (2022) Bayesian hierarchical quantile regression with application to characterizing the immune architecture of lung cancer. *Biometrics*, 1–15. https://doi.org/10.1111/biom.13774