**Received:** Se e be 2, 2024. **Revised:** N e. be 4, 2024. **Accepted:** N e. be 29, 2024

and divergent responses to immunotherapy, even suggesting potential immune evasion in cases of high TMB. These findings question the adequacy of a unidimensional TMB metric and advocate for clonal heterogeneity-adaptive modifications of the biomarker-based framework.

Second, most TMB analyses simplify statistical models by concentrating on a single efficacy endpoint, such as the Keynote-158 trial, which underpinned TMB's FDA approval based solely on response rates [22]. The ICI impact on varied endpoints, such as objective response rate (ORR) and time-to-event (TTE) measures like overall survival (OS) and progression-free survival (PFS), often differs substantially [23–27]. ORR assesses immediate tumor shrinkage, while TTE gauges long-term therapeutic effects. The lack of assessment of survival endpoints inadequately reflects the multifacetedness of ICI treatments [28, 29]. An integrated approach that merges these co-primary endpoints could significantly enhance statistical power and is essential for capturing the full spectrum of responses to advanced cancer therapies, thereby maximizing the utility of TMB in immunotherapy evaluations [30, 31].

Accordingly, we ideally seek to develop robust statistical models that manage multiple co-primary endpoints while capturing interactions among clonal variables. By refining the TMB framework to encompass both clonal and endpoint heterogeneity, we significantly augment the prognosis and tailor ICI decision-making to individual tumor profiles. However, this enhancement leads to the third considerable challenge. Due to substantial costs and potential adverse effects, ICI trials often have limited sample sizes. Statistical inference is hindered by differing levels of data availability across diverse molecular sub-types and treatment responses [32–35]. Traditional methods that pool data using an identical regression model, such as meta-analyses aggregating several published studies, might introduce artifacts from batch effects and tend to yield inaccuracies [7]. While models customized for specific groups struggle with sparse data, complicating parameter estimations [36]. The variability of observational data across different subgroups necessitates the development of flexible, efficient, and scalable statistical models.

Therefore, this work presents a TMB Heterogeneity-Optimized Regression (**THOR**) model, an innovative approach that integrates clonality mutational profiles from immunotherapy cohort-structured patients. Under the generalized linear mixed model (GLMM) framework, **THOR** navigates multiple clinical endpoints and uncovers potential associations through random effects [37]. The heterogeneity-adaptive penalty terms can effectively address multicollinearity within clonal features, facilitating the identification of pivotal variables that enhance efficacy [38], while accommodating small sample sizes by considering coefficient differences across subgroups, thus promoting data sharing and adaptability across varied cohorts. To validate the clinical utility, we assembled 238 samples covering non-small cell lung cancer (NSCLC), melanoma, and nasopharyngeal carcinoma (NPC) from the Second Affiliated Hospital of Xi'an Jiaotong University and Sun Yat-sen University Cancer Center (SYUCC) to form group-structured data. Additionally, 2212 samples, including genomic mutational profiles and clinicopathologic data, were collected from public studies. The effectiveness of **THOR** has been rigorously validated through statistical simulations and clinical experiments, demonstrating exceptional patient stratification and prediction accuracy capabilities. To foster wider adoption and continuous improvement of this modeling technique, we have made the **THOR** freely available on https://github.com/YixuanWang1120/THOR_project.

# Methods

**THOR** is an advanced computational framework that predicts immunotherapy response based on genomic mutational profiles in group-structured data. This model captures endpoint heterogeneity through random effects to illuminate the dynamic relationships. Additionally, **THOR** incorporates a penalty term within its joint likelihood to manage the intricate covariance observed in tumor clonal heterogeneity, along with a fusion operator to account for coefficient differences across specific patient subgroups, significantly enhancing data integration across heterogeneous cohorts. **THOR** improves statistical validity and interpretability by synthesizing multiple endpoints, addressing covariances, and optimizing for limited sample sizes, strengthening data-driven support for immunotherapy decision-making. A flowchart depicting the specification of **THOR** can be found in Fig. 1.

## Endpoint heterogeneity-adaptive model

The extension of GLMM to encapsulate endpoint heterogeneity across $K \in \mathbb{N}$ grouped cohorts delineates a nuanced approach to analyzing complex biomedical data. Each patient within these cohorts harbors genomic mutations derived from $M \in \mathbb{N}$ subclones containing $p \in \mathbb{N}$ features. That is, we consider subgroup-specific regression indexed by $k$.

For each patient $i \in \{1, \ldots, n_k\}$ in subgroup $k$, the response variable $R_i^{(k)}$ indicates tumor remission, classified by the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 [39], where response and non-response are recorded as 1 or 0. The observed event (tumor recurrence, disease progression, or death, etc.) time $T_i^{(k)}$ takes the minimum of the actual event time $T_i^{(k)*}$ and censoring time $C_i^{(k)}$, with an indicator $\delta_i^{(k)} = \mathbb{I}(T_i^{(k)*} \leq C_i^{(k)})$ signifying whether censoring has occurred. We consider the individual feature matrix $\mathbf{X}_i^{(k)} \in \mathbb{R}^{M \times p}$, with each row $\mathbf{X}_{i,m}^{(k)}$ denoting the feature from the $m$th subclone. The integration of binary outcomes, ORR, and continuous survival times, TTE, is given by the parametric functional:

$$\begin{cases} \mathcal{J}\left(R_i^{(k)}\right) = \sum_{m=1}^{M} \mathbf{X}_{i,m}^{(k)} \boldsymbol{\alpha}_{k,m} + \omega_i^{(k)} \\ h_i^{(k)}(t) = h_0(t)\, e^{\left(\sum_{m=1}^{M} \mathbf{X}_{i,m}^{(k)} \boldsymbol{\beta}_{k,m} + \omega_i^{(k)}\right)} \end{cases} \tag{1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent the fixed-effects coefficients; $h(t)$ symbolizes the instantaneous risk; $h_0(t)$ denotes the baseline hazard; $\omega_i^{(k)}$ denotes the patient-specific random effect. ORR and TTE endpoints are interconnected through the shared random effect, $\boldsymbol{\omega}$, which adheres to a zero-mean normal distribution $N(0, \sigma^2)$. This configuration captures the intra-individual variability and the correlation of endpoints, thus maintaining conditional independence given $\boldsymbol{\omega}$. The sophistication of this model is highlighted in the joint likelihood, which aggregates contributions across different patients and endpoints:

$$L(\mathbf{R}, \mathbf{T}, \Delta; \boldsymbol{\Theta}) = \prod_k \prod_i \int_\omega f\left(R_i^{(k)} \mid \omega_i^{(k)}; \boldsymbol{\Theta}\right) f\left(T_i^{(k)}, \delta_i^{(k)} \mid \omega_i^{(k)}; \boldsymbol{\Theta}\right) f\left(\omega_i^{(k)}; \boldsymbol{\Theta}\right) d\omega \tag{2}$$

Figure 1. Flow of **THOR** methodology.

and the specific likelihood contribution for the ith individual is given by:

$$l(\boldsymbol{\Theta}) = \int_{\omega} \left\{ \left[ \frac{e^{\eta}}{1+e^{\eta}} \right]^{R} \left[ \frac{1}{1+e^{\eta}} \right]^{1-R} \left[ h_0(T) e^{\zeta} \right]^{\delta} e \quad \left[ -\mathcal{H}_0(\eta) e^{\zeta} \right] \left[ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\omega^2}{2\sigma^2}} \right] \right\} d\omega \tag{3}$$

where $\mathcal{H}_0(t) = \int_0^t h_0(s)ds$ is the cumulative baseline risk; $\eta$ and $\zeta$ represent the sum of linear predictors over different endpoints, respectively. GLMM typically eliminates the unknown baseline hazards to retain the cancelation property in the parameters estimation. Thus, the partial likelihood for the TTE data is structured as:

$$f_{pl}(\boldsymbol{T}, \Delta) = \left[ \frac{e^{\zeta}}{\sum_{s \in R(t_i)} e^{\zeta_s}} \right]^{\delta} \tag{4}$$

where $R(t)$ is defined as $\{s : T_s \geq t\}$, representing the risk set at time $t-$.

Integration over random effects is challenging and requires numerical integration methods such as Laplace Approximation and Gaussian Hermite Quadrature. Optimization algorithms, notably the Newton–Raphson and the Expectation Maximization algorithm, are then applied.

## Clonal heterogeneity-adaptive penalty

Next, we explore the implementation of clonal heterogeneity-adaptive penalties within the joint likelihood. Tumors feature genetically varied subclones that evolve through cell division. These subclones adapt to the tumor microenvironment, enhancing survival and proliferation and engaging in symbiotic interactions that may modify this environment. Mutational correlations in clonality data can cause considerable covariance, destabilizing parameter estimations and reducing predictive accuracy. Therefore, we apply regularization to improve model output stability and interpretability and analyze complicated biological data more robustly. For illustration, we consider the model formulation under a single cohort ($K = 1$).

LASSO applies an $\ell_1$ penalty to the loss function, boosting model coefficient sparsity. This is especially effective when the feature space has more dimensionality than samples, resulting in zero non-predictive coefficients. The likelihood incorporating the LASSO penalty is given by:

$$\mathcal{L}_p(\boldsymbol{\Theta}) = \quad L(\boldsymbol{\Theta}) - \lambda \left[ \|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 \right] \tag{5}$$

where $\lambda$ is a hyperparameter that controls the shrinkage degree, aiding in model simplification and enhancing interpretability. The selection of $\lambda$ generally requires computationally demanding cross-validation (CV) for meticulous tuning. However, a limitation of LASSO is its potential to omit relevant correlated variables and underestimate non-zero coefficients, introducing bias—particularly problematic in complex tumor analyses.

Ridge regression, an $\ell_2$ variant, does not promote sparsity but offers continuous differentiability and smoothness in optimization, contrasting with the non-convex nature of LASSO. The Ridge penalized likelihood is:

$$\mathcal{L}_p(\boldsymbol{\Theta}) = \quad L(\boldsymbol{\Theta}) - \frac{\lambda^2}{2} \left[ \|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2 \right] \tag{6}$$

With the penalties, parameter estimation via maximum likelihood estimates (MLE) necessitates using iterative algorithms such as coordinate descent or gradient-based optimization methods. Based on the observed dataset $\mathcal{D} = \{\mathbf{R}, \mathbf{T}, \Delta, \mathbf{X}\}$, the process involves a sequence of steps to iteratively estimate the unknown parameters $\boldsymbol{\Theta} = \left[ \boldsymbol{\alpha}_1^{\top}, \ldots, \boldsymbol{\alpha}_M^{\top}, \boldsymbol{\beta}_1^{\top}, \ldots, \boldsymbol{\beta}_M^{\top}, \sigma \right]^{\top}$. The Lasso penalty's non-convex likelihood function is a major computing barrier in

optimization, thus it is updated via a soft-thresholding operation to promote sparsity. When substituting the Lasso penalty with Ridge, the non-convex character of the $\ell_1$ norm is replaced with the $\ell_2$ norm's continuous derivability and smoothness, simplifying the optimization process. Here's a pseudo-code representation providing a structured description of the penalized MLE process for a joint model:

**Algorithm 1** Penalty for tumor clonal heterogeneity

**Input:** Observed dataset $\mathcal{D} = \{\mathbf{R}, \mathbf{T}, \Delta, \mathbf{X}\}$, $\epsilon$, $n_{\text{iter}}$
**Output:** Penalized MLEs, $\hat{\mathbf{\Theta}}_p$
**Step 1: Joint Model for Clinical Endpoints**
$\mathcal{L}_p(\mathbf{\Theta}) = \left(\prod \int\int f(R \mid \omega; \mathbf{\Theta}) f(T, \Delta \mid \omega; \mathbf{\Theta}) f(\omega; \mathbf{\Theta})\, d\omega\right) - \text{penalty}(\mathbf{\Theta})$
**Step 2: Numerical Integration Using Gauss-Hermite Quadrature**
$\int_{-\infty}^{\infty} f(x) e^{-x^2}\, dx \approx \sum_{d=0}^{n} w_d f(x_d)$, where $w_d$ are weights and $x_d$ are nodes of the quadrature.
**Step 3: Iterative Estimation**
Initialize: $iter \leftarrow 0$
**while** $|\mathbf{\Theta}_p^{(j+1)} - \mathbf{\Theta}_p^{(j)}| > \epsilon$ and $iter < n_{\text{iter}}$ **do**
$\mathbf{\Theta}_p^{\text{temp}} \leftarrow \text{DescentUpdate}(\mathcal{L}_p(\mathbf{\Theta}_p^{(j)}))$
$\mathbf{\Theta}_p^{(j+1)} \leftarrow \text{sgn}(\mathbf{\Theta}_p^{\text{temp}}) \cdot \quad \text{a} \quad (|\mathbf{\Theta}_p^{\text{temp}}| - \lambda, 0)$
$iter \leftarrow iter + 1$
**end while**

## Subgroup heterogeneity-adaptive fusion

Conventional immunotherapy study analyses assume observational sample consistency and use a singular model for inferential statistics. Valid conclusions depend on this homogeneity assumption. However, the intrinsic characteristics of immunotherapy, such as high costs and notable side effects, often result in datasets comprising heterogeneous cohorts with limited sample sizes. These practical constraints necessitate a more nuanced approach that accounts for subgroup diversity.

Therefore, we introduce the joint lasso, an innovative extension of the penalized joint likelihood framework designed for cross-sub-cohort data fusion. In contrast to simple pooling, our approach allows heterogeneous subgroups to have distinct sparsity patterns and regression coefficients, utilizing commonalities and accommodating differences. This method effectively balances the need for subgroup-specific modeling with the advantages of borrowing strength across similar subgroups. A penalization operator is used to integrate coefficient discrepancies between sub-cohorts into the likelihood:

$$\mathcal{L}_{pf}(\mathbf{\Theta}) = \sum_{k=1}^{K} \Bigg\{ \quad L_k(\mathbf{\Theta}) - \frac{\lambda^2}{2}\left[\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2\right] + \gamma \sum_{k'>k} \left[\tau_{k,k'} \left(\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k'}\|_2^2 + \|\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k'}\|_2^2\right)\right] \Bigg\} \tag{7}$$

where $\lambda$, $\gamma$, $\tau$ regulate penalized shrinkage, and $\tau_{k,k'}$ controls the degree of similarity for certain subgroup pairs.

By default, fusion parameters $\tau$ can be set uniformly to 1, proposing a baseline unweighted fusion across all subgroups. However, $\tau$ is dynamically modified to reflect a weighted fusion to account for subgroup variability. While CV is commonly employed to fine-tune, it can prove burdensome. Alternatively, a more interpretable method involves setting $\tau_{k,k'}$ based on a feature-based distance metric $d(k, k')$, enhancing fusion among subgroups that exhibit more remarkable feature similarity. This method promotes coherence and efficiency in the model by directly linking parameter adjustments to measurable subgroup similarities.

Choosing an efficient distance metric is critical for clonal mutation features. Euclidean distance can be employed to gauge superficial differences between the means of sample characteristics within each subgroup. Mahalanobis distance, incorporating the covariance, can reflect potential correlations among subclones, providing deeper insight. Here, the symmetric Kullback–Leibler (KL) divergence is introduced to compare subgroup feature distributions. Its capacity to balance mutational feature probability distribution disparities inspired this choice:

$$d\left(k, k'\right) = \frac{1}{2} \sum_{m} \left(D_{KL}\left(\hat{p}_{k,m} \| \hat{p}_{k',m}\right) + D_{KL}\left(\hat{p}_{k',m} \| \hat{p}_{k,m}\right)\right)$$
$$D_{KL}(p \| q) = \sum_{x} p(x) \qquad \frac{p(x)}{q(x)} \tag{8}$$

where $\hat{p}_{k,m}$ and $\hat{p}_{k',m}$ denote the estimated distributions of mutational features for the $m$th subclone in subgroup $k$ and $k'$, whose distributions are typically modeled as multivariate normal.

With the feature-based distance metric established, the fusion parameter $\tau_{k,k'}$ is set according to:

$$\tau_{k,k'} = 1 - d\left(k, k'\right) / d \quad \text{a} \tag{9}$$

where $d$ a represents the maximum distance observed among all subgroup pairs. This parameterization adjusts fusion parameters proportionally based on subgroup similarities and differences. Here's a pseudo-code representation, providing a structured and precise description of the fusion process for **THOR**:

**Algorithm 2** Fusion MLE across subgroup heterogeneity

**Input:** Observed dataset for $K$ subgroups $\mathcal{D}_k = \{\mathbf{R}_k, \mathbf{T}_k, \Delta_k, \mathbf{X}_k\}$, $\epsilon$ and $n_{\text{iter}}$
**Output:** Fusion MLEs, $\hat{\mathbf{\Theta}}_{pf}$
**Step 1: Fusion likelihood across subgroups**
$\mathcal{L}_{pf}(\mathbf{\Theta}) = \sum_{k=1}^{K} \Big\{ \quad L_k(\mathbf{\Theta}) - \frac{\lambda^2}{2}\left[\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_2^2\right] + \gamma \sum_{k'>k} \left[\tau_{k,k'} \left(\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k'}\|_2^2 + \|\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k'}\|_2^2\right)\right] \Big\}$
**Step 2: Set fusion parameters $\tau_{k,k'}$ based on distance metric**
$\tau_{k,k'} = 1 - d\left(k, k'\right) / d \quad \text{a}$
**Step 3: Iterative Estimation**
Initialize: $iter \leftarrow 0$
**while** $|\mathbf{\Theta}_{pf}^{(j+1)} - \mathbf{\Theta}_{pf}^{(j)}| > \epsilon$ and $iter < n_{\text{iter}}$ **do**
$\mathbf{\Theta}_{pf}^{(j+1)} \leftarrow \mathbf{\Theta}_{pf}^{(j)} - H_{pf}^{-1}(\mathbf{\Theta}_{pf}) \cdot \nabla \mathcal{L}_{pf}$
$iter \leftarrow iter + 1$
**end while**

## Statistical analyses

We conducted all computational analyses using Python (version 3.10). The **THOR** model used existing libraries to facilitate data processing, model fitting, and evaluation. We employed NumPy (version 1.26.4) and Pandas (version 2.2.2) for numerical calculations and data handling. Optimization routines and statistical functions were implemented using SciPy (version 1.12.0). Machine learning algorithms and model evaluation metrics were managed using scikit-learn (version 1.2.2), while XGBoost (version 2.0) was employed for comparative analyses. Survival analyses and statistical tests, such as the log-rank test, were conducted using the Lifelines (version 0.27.4).

The computational framework of the **THOR** model handles group-structured data and captures TMB heterogeneity across

Figure 2. Comparative analysis of MSE in parameter estimation. (**A**) MSE variation by model and random effect intensity: box plots compare MSEs across three models, **THOR**, Logistic Regression, and Cox Proportional Hazards Regression, under varying levels of random effect intensity. The plots illustrate that **THOR** consistently maintains lower MSE in rising random effect scenarios, demonstrating its robustness in handling uncertainty. (**B**) MSE across different regularizations: the box plot shows MSE outcomes for four regularization strategies: non-penalty, Lasso, Ridge, and elastic net. The bar chart segments MSE contributions by relevant and non-relevant features, highlighting the effectiveness of regularizations in reducing error, specifically through feature relevance. (**C**) Comparison of **THOR** and separate GLMM: the MSE comparisons between **THOR** and a separate GLMM are depicted. Box plots delineate the lower MSE achieved by **THOR**, emphasizing its superior accuracy in parameter estimation compared to traditional GLMM approaches in separate analyses.

baseline model exhibits a wide distribution of MSE, indicative of increased volatility and a susceptibility to overfitting. In contrast, implementing Lasso and Ridge regularizations substantially narrows this MSE distribution, reflecting improved performance. The Elastic Net strategy, merging the strengths of Lasso and Ridge, attains a median MSE that aligns closely with these two methods, ensuring balanced efficacy across diverse feature types and establishing a holistic approach to regularization.

Figure 2C visually compares MSE in parameter estimation between **THOR** and the Separate GLMM analysis. The Separate GLMM involves conducting separate analyses of different subgroups of data without accounting for potential correlations between groups. **THOR** achieves a lower median MSE of approximately 0.03, signaling higher precision and diminished error variability compared to separate regressions, which display a median MSE of roughly 0.042. The broader interquartile range and heightened error variability associated with the Separate GLMM analysis suggest its reduced effectiveness in capturing complex data features, particularly in datasets with significant subgroup heterogeneity.

## Clinical cohorts application analysis

### *Participants and data production*

In this research, we performed a retrospective analysis on group-structured data consisting of 238 patients who received ICI monotherapy. Participants included 64 patients diagnosed with recurrent/metastatic nasopharyngeal carcinoma (R/M NPC) and 73 patients with NSCLC at SYUCC, as well as 75 patients with NSCLC and 26 patients with melanoma from the Second Affiliated Hospital of Xi'an Jiaotong University. Patients afflicted with R/M NPC were enrolled in two Phase I clinical trials (NCT02721589 and NCT02593786). Protocols for NSCLC patients at SYUCC, including dosage escalation and expansion phases, are detailed in the references [40–42]. Additionally, 101 patients from the Second Affiliated Hospital of Xi'an Jiaotong University underwent analogous treatment protocols and were sequenced at the Geneplus-Beijing Institute. Patient specifics are delineated in Supplementary Table S2, with extended information on patient enrollment, library preparation, sequencing, and bioinformatics procedures available in the Supplementary Materials. This synthesis of data from multiple clinical cohorts facilitates an evaluation of the prognostic significance of clonal mutation characterization, focusing on clinical endpoints such as ORR and PFS. Four related but distinct experimental subgroups were naturally identified based on disease type or treatment, specifically Geneplus_Lung, Geneplus_SKCM, SYUCC_Lung, and SYUCC_NPC.

To further verify the model's clinical applications, we augmented the dataset by collecting a total of 2212 samples, which included both genomic mutational profiles and clinicopathologic data [3, 43–51]. Based on the study source and disease type, the validation assembly incorporated NSCLC cohorts with 467 patients structured into four subgroups, Melanoma cohorts with 720 patients structured into six subgroups, and Pan-caner cohorts with 1025 patients structured into nine subgroups. Fig. 3A and B illustrates the proportionate distribution of individuals within each cohort and details the distribution of cancer types across all clinical cohorts. The entirety of the clinical cohorts and the corresponding grouping structure is briefly detailed in Table 1.

Figure 3. Analysis of patient cohorts and genomic mutation profiles. (**A**) Sample size distribution across cohorts. (**B**) Cancer type distribution. (**C**) Distribution of CCF: Violin plots show the distribution of CCF across four experimental cohorts, providing insights into the clonal heterogeneity. (**D**) CCF heatmap across genes and cohorts: the heatmap details the CCF for various mutations across genes within SYUCC experimental cohorts, with the color gradient representing the extent of clonality.

Table 1. Summary of study cohort information.

| Subgroup ID | Cohort overview | Study source | Num. | Cancer type |
|---|---|---|---|---|
| 1 | **Experiment Cohorts** | Geneplus_Lung | 75 | Non-small Cell Lung Cancer |
| 2 | | Geneplus_SKCM | 26 | Melanoma |
| 3 | | SYUCC_Lung | 73 | Non-small Cell Lung Cancer |
| 4 | | SYUCC_NPC | 64 | Nasopharyngeal Carcinoma |
| 1 | **NSCLC Cohorts** | Rizvi, Science | 16 | Non-small Cell Lung Cancer |
| 2 | | Hellmann, Cancer Cell | 68 | Non-small Cell Lung Cancer |
| 3 | | Rizvi, JCO | 227 | Non-small Cell Lung Cancer |
| 4 | | Samstein, Nat Genet Lung | 156 | Non-small Cell Lung Cancer |
| 1 | **Melanoma Cohorts** | Hugo, Cell | 37 | Metastatic Melanoma |
| 2 | | Liu, Nat Med | 140 | Metastatic Melanoma |
| 3 | | Riaz, Cell | 70 | Metastatic Melanoma |
| 4 | | Roh, Sci Transl Med | 48 | Metastatic Melanoma |
| 5 | | Van Allen, Science | 105 | Metastatic Melanoma |
| 6 | | Samstein, Nat Genet SKCM | 320 | Metastatic Melanoma |
| 1 | **Pan-cancer Cohorts** | Miao, Science | 35 | Renal Cell Carcinoma |
| 2 | | Samstein, Nat Genet KIRC | 151 | Renal Cell Carcinoma |
| 3 | | Samstein, Nat Genet BLCA | 215 | Bladder Cancer |
| 4 | | Samstein, Nat Genet BRCA | 44 | Breast Cancer |
| 5 | | Samstein, Nat Genet Unknown | 110 | Unknown Primary |
| 6 | | Samstein, Nat Genet COAD | 126 | Colorectal Cancer |
| 7 | | Samstein, Nat Genet ESCA | 117 | Esophagogastric Cancer |
| 8 | | Samstein, Nat Genet GBMLGG | 139 | Glioma |
| 9 | | Samstein, Nat Genet HNSC | 88 | Head and Neck Cancer |

The datasets we examined exhibited significant variations in patient numbers across different groups, which could compromise statistical robustness. For example, a smaller cohort from Rizvi et al. [43] included only 16 patients, potentially introducing significant bias and error if analyzed in isolation. Similar disparities are evident in other datasets. To circumvent these limitations, we used a fusion framework to improve statistical validity and allow even smaller subgroups to contribute effectively to the inference.

Tumor clonal heterogeneity not only charts the evolutionary trajectory of tumors but also influences patient responses to immunotherapy. The PyClone algorithm [52] was utilized to accurately identify and quantify clonal mutation signatures, revealing potential relationships with immunotherapy outcomes. Below is a detailed description of our bioinformatic pipeline:

- Collection: Raw sequencing data in FASTQ format were sourced from corresponding platforms, with each dataset undergoing rigorous checks for completeness and potential corruption to ensure data integrity.
- Quality control: The FastQC was utilized to control the quality of each sample, scrutinizing reports for low sequencing quality, high contamination levels, or excessive adapters. Following this analysis, data cleanup includes quality filtering, removal of adapters via Cutadapt, and elimination of low-quality reads, short reads, and duplicates to ensure accuracy in subsequent analyses.
- Alignment: Cleaned reads were aligned to the reference genome GRCh38 using BWA-MEM, generating BAM files for detailed examination.
- Variant calling: Variant calling was executed using Mutect2 for single nucleotide variants (SNVs) and small insertions/deletions (Indels), along with CNVkit for copy number variation (CNV) analysis.
- Clonal analysis: Input files, including variant data, allele frequencies, and coverage depth, were prepared for PyClone, facilitating the inference of clonal populations and their proportions.

Our bioinformatics pipeline enables precise tumor heterogeneity characterization, as illustrated in Fig. 3C and D. This figure shows the distribution of the clonal cell fraction (CCF) across experimental cohorts, allowing a visual comparison of tumor cellular compositions and mutational profiles across diverse patient groups. The comprehensive presentation allows for a straightforward comparison of mutational profiles across diverse patient groups, providing more profound insights into the molecular mechanisms underlying tumor biology.

### *Effectiveness of THOR in group-structured data*

We initiated our investigation by validating the clinical efficacy of **THOR** against established predictive models—Logistic Regression, XGBoost, and NN—utilizing two distinct analytic strategies: pooled analysis and separate analysis. These models were assessed for their ability to predict clinical outcomes based on genomic mutational profiles and clinicopathologic data. **THOR** leverages a fusion penalty to facilitate information exchange across tumor clones and patient subgroups, effectively identifying both shared and subgroup-specific patterns. Logistic Regression, a traditional classification approach widely used in ICI prediction [53]; XGBoost, an ensemble learning method based on gradient boosting; and NNs, a multilayer perceptron classifier, served as comparative benchmarks.

Our analysis strategies diverged as follows: pooled analysis amalgamated all subgroup data into a single dataset, training one model across the board, thus neglecting subgroup distinctions. Separate analysis treated each subgroup as an independent dataset, training dedicated models to enhance specificity without sharing information between subgroups, except in the case of **THOR**. Performance was rigorously evaluated using stratified K-fold cross-validation (CV) to preserve subgroup distribution. Metrics including log-loss, accuracy, and area under the curve (AUC) were calculated, and receiver operating characteristic (ROC) curves were visualized to compare model performance.

The findings in Fig. 4 and Table 2 first illustrate **THOR**'s superior performance across the experimental cohorts.

Figure 4A presents the ROC curves under pooled analysis, where our model achieves a superior AUC of 0.77, outperforming logistic (AUC = 0.63), XGBoost (AUC = 0.71), and NN (AUC = 0.75). The pooled analysis approach may obscure subgroup-specific patterns crucial for precise predictions, yet **THOR** successfully captures these patterns, leading to superior predictive performance.

Figure 4B depicts **THOR**'s consistent superiority under separate analysis, outperforming logistic (AUC = 0.72), XGBoost (AUC = 0.69), and NN (AUC = 0.72). This underscores **THOR**'s effectiveness in leveraging subgroup-specific information to enhance predictive accuracy. The experimental cohorts consisted of patient data from four distinct subgroups, and response predictions for each cohort are illustrated in Supplementary Fig. S1. **THOR**'s ability to share information across subgroups enhances performance, particularly in data-limited subgroups, whereas the comparison models are prone to overfitting or underfitting due to insufficient subgroup data.

The Kaplan-Meier survival curves in Fig. 4C and D compare the PFS for patients categorized into high- and low-risk groups based on **THOR**-calculated hazards and Cox-calculated hazards. The log-rank test results reveal significant differences between these groups, illustrating **THOR**'s ability to utilize subgroup-specific information and individual genetic markers effectively.

Table 2 quantitatively reinforces **THOR**'s prognostic precision compared to traditional markers such as TMB, emphasizing its advanced capability in prognostic categorization. Within the SYUCC&Geneplus cohorts, **THOR** consistently outstripped its counterparts, achieving a lower log-loss (0.4848) and a higher accuracy rate (76.36%) across both analysis strategies. A notable decline in performance metrics was observed when switching from multi-dimensional clonal mutation features to a unidimensional TMB marker, validating our hypothesis that integrating clonal heterogeneity significantly enhances immunotherapy decision-making, thus advocating for **THOR**'s innovative approach.

This comprehensive evaluation not only fortifies **THOR**'s applicability in clinical predictions but also highlights its potential to significantly refine immunotherapy strategies through detailed clonal analysis, offering a substantial enhancement over existing models.

### *Extended validation of THOR across diverse cancer types*

Further validation of **THOR**'s efficacy was conducted across multiple cancer settings, including NSCLC, melanoma, and pan-cancer cohorts. The results, summarized in Fig. 5 and Table 2, consistently demonstrate superior performance compared to alternative models.

In the NSCLC cohorts (Fig. 5A), the ROC curves from both pooled and separate analyses indicate that **THOR** consistently

Figure 4. Comparative prognostic analysis across experimental cohorts using **THOR**, Logistic Regression, XGBoost, and NN under pooled and separate analysis. (**A**) ROC curves for group-structured cohorts under pooled analysis (5-Fold CV): These curves display the diagnostic accuracy of the **THOR** model

Table 2. Quantitative performance metrics of THOR versus Logistic, XGBoost and nerual network (NN) under pooled and separate analysis for different predictive features.

| Cohort type | Metric & model | THOR | Pooled logistic | Pooled XGBoost | Pooled NN |
|---|---|---|---|---|---|
| Experimental | Log-Loss | **0.4848** | 0.5426 | 0.5133 | 0.5445 |
| | Accuracy | **76.36%** | 75.96% | 75.98% | 75.96% |
| NSCLC | Log-Loss | **0.5854** | 0.6316 | 0.6465 | 0.5917 |
| | Accuracy | **69.47%** | 65.73% | 63.79% | 66.56% |
| Melanoma | Log-Loss | **0.6127** | 0.6758 | 0.6700 | 0.6800 |
| | Accuracy | **66.17%** | 57.32% | 58.79% | 58.34% |
| Pan-cancer | Log-Loss | **0.6174** | 0.6608 | 0.6538 | 0.6359 |
| | Accuracy | 62.11% | 59.17% | 63.22% | **64.54%** |
| **Cohort type** | **Metric & model** | **Separate logistic** | **Separate XGBoost** | **Separate NN** | |
| Experimental | Log-Loss | 0.663 | 0.5093 | 0.4989 | |
| | Accuracy | 75.94% | 75.51% | 77.01% | |
| NSCLC | Log-Loss | 0.6351 | 0.6589 | 0.6033 | |
| | Accuracy | 64.86% | 61.83% | 68.99% | |
| Melanoma | Log-Loss | 0.6611 | 0.6767 | 0.6442 | |
| | Accuracy | 65.28% | 63.36% | 65.88% | |
| Pan-cancer | Log-Loss | 0.6691 | 0.6372 | 0.6415 | |
| | Accuracy | 65.04% | 63.02% | 65.55% | |
| **Cohort type** | **Metric & model** | **THOR for TMB** | **Pooled logistic for TMB** | **Pooled XGBoost for TMB** | **Pooled NN for TMB** |
| Experimental | Log-Loss | 0.5904 | 0.5452 | 0.5779 | 0.5441 |
| | Accuracy | 75.53% | 75.96% | 75.12% | 75.96% |
| NSCLC | Log-Loss | 0.6954 | 0.6434 | 0.6547 | 0.6392 |
| | Accuracy | 61.85% | 65.73% | 63.99% | 65.94% |
| Melanoma | Log-Loss | 0.6855 | 0.6858 | 0.6917 | 0.6829 |
| | Accuracy | 55.39% | 55.24% | 58.49% | 55.24% |
| Pan-cancer | Log-Loss | 0.6814 | 0.676 | 0.6655 | 0.6672 |
| | Accuracy | 63.22% | 59.17% | 61.80% | 60.89% |

in clinical decision-making. The consistent outperformance in subgroup-specific analyses and broader cohort validations highlights its potential as an essential tool in precision oncology.

### *Overall impact of clonal composition on clinical endpoints*

The four subplots in Fig. 6 represent the clonal tumor mutation burden (cTMB) coefficients for various groups within different cancer cohorts: experimental, NSCLC, melanoma, and pan-cancer. The coefficients quantify the effect of cTMB on two clinical outcomes: ORR and TTE.

In the experimental cohort (Fig. 6A), Clone 1, the master clone with the largest cancer cell fraction (CCF), has the highest absolute coefficient values for ORR and TTE across multiple groups. This emphasizes the major clone's dominance in treatment response and survival. Dashed lines show cTMB's effect on TTE, with negative coefficients showing that larger levels reduce hazard and improve survival. The stronger effect on TTE, relative to ORR, suggests the greater impact of cTMB on long-term survival compared to short-term treatment response. Negative effects observed for certain groups (e.g. Group 3) on ORR also indicate that specific clones can adversely impact ICI outcomes, highlighting the importance of clonal composition in predicting both response and progression.

In the NSCLC cohorts (Fig. 6B), despite variations observed between different groups, the data further corroborate the trend where Clone 1 consistently exerts the greatest influence on clinical outcomes. This suggests that in NSCLC, the primary clone maintains its dominant role, while subclonal contributions are secondary. This trend underscores the importance of the master clone in shaping treatment response and survival trajectories, reaffirming a pattern of clonal dominance as reported previously [19].

The melanoma cohorts (Fig. 6C) deviate significantly from the other cohorts, with the effect of the primary clone (Clone 1) being diluted by substantial contributions from subclones, particularly Clone 2 and Clone 3. This dynamic may partly explain why TMB is less reliable as a predictive marker for melanoma compared to other cancer types. Melanoma tumors exhibit a complex subclonal structure where secondary clones play a substantial role, thereby diminishing the predictive influence of the primary clonal mutation burden. This observation underscores the necessity for more clonality-aware metrics to accurately predict outcomes in melanoma, as traditional TMB may fail to fully capture the intricate subclonal interactions and their impact on patient prognosis.

Although subclones do contribute to clinical outcomes, their effects are less pronounced compared to those of the primary clone. In the pan-cancer analysis (Fig. 6D), the consistency of this trend further emphasizes the significance of the clones with larger CCF in overall prognostic assessments, though it also implies that cancer-specific nuances may sometimes be obscured by generalized analytical approaches.

The visualized coefficients underscore the importance of cTMB in predicting clinical outcomes such as ORR and TTE across different cancer types and subgroups. The variability across subgroups suggests that a one-size-fits-all prediction or treatment strategy could be ineffective or even detrimental. Understanding the specific contributions of different clones to ORR and TTE facilitates more informed clinical decision-making, potentially enhancing both initial treatment responses and long-term outcomes. **THOR**'s

Figure 5. Comparative prognostic analysis across validation cohorts using **THOR**, Logistic Regression, XGBoost, and NN under pooled and separate analysis. (**A**) NSCLC cohorts: ROC curves indicate that **THOR** achieves the highest AUC (0.70), surpassing all other models and demonstrating superior predictive accuracy. Kaplan–Meier survival curves (PFS) show significant separation between high and low-risk groups under **THOR**, with a log-rank test *P*-value of 0.0003, compared to 0.0072 under pooled analysis. (**B**) Melanoma cohorts: ROC curves show **THOR** achieves the highest AUC (0.73), outperforming all alternative models. Kaplan–Meier survival curves for PFS further affirm **THOR**'s superior predictive precision, achieving a log-rank test *P*-value of 0.0000, underscoring its strong prognostic capability in melanoma cohorts. (**C**) Pan-cancer cohorts: ROC curves demonstrate **THOR**'s superior prognostic accuracy, with an AUC of 0.70, outperforming other models across diverse cancer types. Kaplan-Meier survival curves (OS) demonstrate the effectiveness of **THOR** in universally applicable cancer prognostics, with consistently low log-rank test *P*-values, indicating significant differentiation between risk groups.

ability to capture these variations across subgroups and clones demonstrates its utility in providing tailored risk stratification, surpassing the capacity of traditional models. This capability is particularly critical in precision oncology, where therapeutic decisions must consider subtle differences across diverse patient subgroups and cancer types.

## Discussion

This research has presented **THOR**, a heterogeneity-optimized regression model designed to predict immunotherapy responses by analyzing clonal genomic features within group-structured data. The model effectively addresses challenges related to integrating multi-endpoint data, managing high covariance levels, and limited sample sizes, thereby strengthening statistical inference for personalized immunotherapy. By leveraging a GLMM framework, **THOR** unites various clinical outcomes, such as ORR and PFS, into a cohesive analytical approach.

**THOR** distinguishes itself by mitigating multicollinearity among clonal mutations using a penalized likelihood approach that emphasizes crucial variables. This method, combined with a fusion strategy across data subgroups, preserves model specificity while enhancing adaptability and predictive accuracy through efficient information sharing. Simulations underscore the model's proficiency in parameter estimation and its ability to handle nonconvex optimizations, which are often encountered in complex models. When applied to a clinical dataset of nearly a thousand patients with conditions such as NSCLC and melanoma, **THOR** demonstrated enhanced risk stratification and superior predictive performance compared to conventional methods.

Despite these promising outcomes, several limitations warrant discussion. First, **THOR** relies heavily on the availability of high-quality genomic and clinical data, which are not consistently accessible across different healthcare systems and research environments. This reliance could hinder the model's widespread applicability, especially in scenarios where genomic sequencing is incomplete or unavailable, limiting its use to well-funded research settings or specialized cancer centers. Additionally, **THOR** requires substantial computational resources, particularly for model fitting, which involves penalized likelihood estimation and complex optimization procedures. These computational demands may restrict its feasibility in environments with limited computing infrastructure.

The current implementation of **THOR** utilizes GLMMs, which, although powerful, are computationally intensive and present challenges related to nonconvex optimization. Future work will explore reducing these computational barriers, potentially incorporating efficient approximation algorithms, or leveraging distributed computing frameworks to enhance scalability. Addressing these challenges is crucial to making the model more accessible for broader clinical application.

To address these limitations, future work will focus on refining **THOR** by incorporating more comprehensive and diverse datasets, thereby broadening the model's scope. This effort includes

Figure 6. Analysis of cTMB coefficients by clone and group across different cancer cohorts. Figure depicts the estimated coefficients for cTMB on clinical endpoints, represented as the effect on ORR and TTE, across various clones and patient subgroups in different cancer types. The coefficient values are derived from **THOR** and visually depict how clonal structures influence distinct clinical outcomes. (**A**) cTMB coefficients for the experimental cohorts. The panel illustrates coefficient estimates for four subgroups across five distinct clones (Clone 1 to Clone 5). The solid lines indicate the estimated effect of cTMB on ORR, while the dashed lines represent the effect on TTE. The primary clone (Clone 1) consistently shows the largest coefficient values in absolute terms, particularly on TTE, suggesting that higher cTMB levels reduce the survival hazard and improve prognosis. (**B**) cTMB coefficients for the NSCLC cohorts. This panel shows the effect of cTMB across four subgroups. Clone 1 generally retains the largest effect on both ORR and TTE. The influence of subclones, especially Clone 3 in Group 2, is also evident, although their effect does not exceed that of the master clone. (**C**) cTMB coefficients for the melanoma cohorts. The panel illustrates coefficient values across six groups and five clones. Notably, the influence of the primary clone is diluted by the substantial contributions from subclones (Clone 2 and Clone 3). Dashed lines, representing the cTMB effect on TTE, are largely negative across the subclones, implying that higher cTMB levels are associated with reduced hazard and improved survival. (**D**) cTMB coefficients for the pan-cancer cohorts. Coefficients are shown for nine groups across four clones. Clone 1 consistently displays the highest coefficient values across both endpoints. The dashed lines for TTE indicate a negative relationship between cTMB and survival hazard, showing a generally consistent effect.

acquiring richer genomic and clinical data, extending the analysis to additional clinical endpoints—such as adverse event rates—and integrating a broader range of gene mutation signatures across various tumor types. We also plan to explore advanced machine learning techniques, including deep learning-based feature extraction, to effectively manage the inherent complexity and high dimensionality of genomic data.

## Conclusion

**THOR** represents a significant advancement in predictive oncology, with substantial potential to improve patient-specific outcomes in cancer immunotherapy. By effectively addressing the challenges of endpoint integration, subgroup fusion, and high-dimensional data, **THOR** has demonstrated its capacity for personalized response prediction. Future research aimed at enhancing clinical applicability and computational efficiency is expected to significantly advance the field toward more precise and individualized therapeutic strategies, ultimately improving patient care and outcomes in oncology.

**Key Points**

- We introduced a heterogeneity-optimaized regression (**THOR**), an advanced computational framework that predicts immunotherapy response based on genomic mutational profiles in group-structured data.
- **THOR** enhances the predictive capabilities of TMB by integrating tumor clonality and clinical co-primary endpoints, further augmented by fusion techniques across subgroups, improving the statistical power and prompting data sharing.
- **THOR**'s effectiveness is demonstrated through simulations and clinical application on a cohort of 238

cancer patients, supplemented by data from 2212 patients, showing significant improvements in patient stratification and prognostic accuracy.

## Acknowledgements

We thank the patients and their families for participation in the study.

## Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Conflict of interest: Author G.Y. is employed by Geneplus-Beijing Institute. Authors C.Z. and W.X. are employed by Nanjing Geneseeq Technology Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential competing interest.

## Funding

The research presented here was supported by the National Natural Science Foundation of China (No.62302215), and the Postdoctoral Fellowship Program of CPSF (No.GZC20233494).

## Data availability

The code and data of **THOR** are available at https://github.com/YixuanWang1120/THOR_project.

## Author contributions

W.Y., G.Y., W.J., and S.X. designed the research; W.Y., L.X., and W.J. developed the mathematical model; W.Y., G.Y., L.Y., and Y.S. performed the research and numerical simulations; W.Y., G.Y., C.Z., W.X., W.Q., L.J., and Y.S. collected and analyzed the data; W.Y., Z.J., W.J., Y.S., and S.X. wrote the paper.

## References

1. Morad G, Helmink BA, Sharma P. *et al.* Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* 2021;**184**:5309–37. https://doi.org/10.1016/j.cell.2021.09.020.
2. Schmid P, Cortes J, Dent R. *et al.* Event-free survival with pembrolizumab in early triple-negative breast cancer. *N Engl J Med* 2022;**386**:556–67. https://doi.org/10.1056/NEJMoa2112651.
3. Samstein RM, Lee C-H, Shoushtari AN. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* 2019;**51**:202–6. https://doi.org/10.1038/s41588-018-0312-8.
4. Valero C, Lee M, Hoen D. *et al.* Pretreatment neutrophil-to-lymphocyte ratio and mutational burden as biomarkers of tumor response to immune checkpoint inhibitors. *Nat Commun* 2021;**12**:729. https://doi.org/10.1038/s41467-021-20935-9.
5. Goodman AM, Kato S, Bazhenova L. *et al.* Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther* 2017;**16**:2598–608. https://doi.org/10.1158/1535-7163.MCT-17-0386.
6. Subbiah V, Solit DB, Chan TA. *et al.* The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB)≥ 10: a decision centered on empowering patients and their physicians. *Ann Oncol* 2020;**31**:1115–8. https://doi.org/10.1016/j.annonc.2020.07.002.
7. Budczies J, Kazdal D, Menzel M. *et al.* Tumour mutational burden: clinical utility, challenges and emerging improvements. *Nat Rev Clin Oncol* 2024;**21**:725–42. https://doi.org/10.1038/s41571-024-00932-9.
8. Jardim DL, Goodman A, De Melo Gagliato D. *et al.* The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* 2021;**39**:154–73. https://doi.org/10.1016/j.ccell.2020.10.001.
9. McGrail DJ, Pilié PG, Rashid NU. *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol* 2021;**32**:661–72. https://doi.org/10.1016/j.annonc.2021.02.006.
10. Anagnostou V, Bardelli A, Chan TA. *et al.* The status of tumor mutational burden and immunotherapy. *Nat Cancer* 2022;**3**:652–6. https://doi.org/10.1038/s43018-022-00382-1.
11. Li X, He Y, Jiashuo W. *et al.* A novel pathway mutation perturbation score predicts the clinical outcomes of immunotherapy. *Brief Bioinform* 2022;**23**:bbac360. https://doi.org/10.1093/bib/bbac360.
12. Li X, He Y, Jiang Y. *et al.* PathwayTMB: a pathway-based tumor mutational burden analysis method for predicting the clinical outcome of cancer immunotherapy. *Mol Ther Nucleic Acids* 2023;**34**:102026. https://doi.org/10.1016/j.omtn.2023.09.003.
13. Niknafs N, Balan A, Cherry C. *et al.* Persistent mutation burden drives sustained anti-tumor immune responses in human cancers. *Cancer Res* 2023;**83**:440–449. https://doi.org/10.1158/1538-7445.AM2023-1393.
14. Leong TL, Aloe C, Aujla S. *et al.* Heterogeneity of tumour mutational burden in metastatic NSCLC demonstrated by endobronchial ultrasound sampling. *Front Oncol* 2023;**13**:1150349. https://doi.org/10.3389/fonc.2023.1150349.
15. Haragan A, Field JK, Davies MPA. *et al.* Heterogeneity of PD-L1 expression in non-small cell lung cancer: implications for specimen sampling in predicting treatment response. *Lung Cancer* 2019;**134**:79–84. https://doi.org/10.1016/j.lungcan.2019.06.005.
16. Kazdal D, Endris V, Allgäuer M. *et al.* Spatial and temporal heterogeneity of panel-based tumor mutational burden in pulmonary adenocarcinoma: separating biology from technical artifacts. *J Thorac Oncol* 2019;**14**:1935–47. https://doi.org/10.1016/j.jtho.2019.07.006.
17. Leong TL, Gayevskiy V, Steinfort DP. *et al.* Deep multi-region whole-genome sequencing reveals heterogeneity and gene-by-environment interactions in treatment-naive, metastatic lung cancer. *Oncogene* 2019;**38**:1661–75. https://doi.org/10.1038/s41388-018-0536-1.
18. Mandal R, Samstein RM, Lee K-W. *et al.* Genetic diversity of tumors with mismatch repair deficiency influences anti–PD-1 immunotherapy response. *Science* 2019;**364**:485–91. https://doi.org/10.1126/science.aau0447.
19. Litchfield K, Reading JL, Puttick C. *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 2021;**184**:596–614.e14. https://doi.org/10.1016/j.cell.2021.01.002.
20. Wolf Y, Bartok O, Patkar S. *et al.* UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell* 2019;**179**:219–235.e21. https://doi.org/10.1016/j.cell.2019.08.032.
21. Zhu M, Kim J, Deng Q. *et al.* Loss of p53 and mutational heterogeneity drives immune resistance in an autochthonous

mouse lung cancer model with high tumor mutational burden. *Cancer Cell* 2023;**41**:1731–1748.e8. https://doi.org/10.1016/j.ccell.2023.09.006.
22. Prasad V, Addeo A. The FDA approval of pembrolizumab for patients with TMB> 10 mut/Mb: was it a wise decision? No. *Ann Oncol* 2020;**31**:1112–4. https://doi.org/10.1016/j.annonc.2020.07.001.
23. Cao D, Huilin X, Ximing X. *et al.* High tumor mutation burden predicts better efficacy of immunotherapy: a pooled analysis of 103078 cancer patients. *Onco Targets Ther* 2019;**8**:e1629258. https://doi.org/10.1080/2162402X.2019.1629258.
24. Valero C, Lee M, Hoen D. *et al.* The association between tumor mutational burden and prognosis is dependent on treatment context. *Nat Genet* 2021;**53**:11–5. https://doi.org/10.1038/s41588-020-00752-4.
25. Tan AC, Bagley SJ, Wen PY. *et al.* Systematic review of combinations of targeted or immunotherapy in advanced solid tumors. *J Immunother Cancer* 2021;**9**:e002459. https://doi.org/10.1136/jitc-2021-002459.
26. Ritchie G, Gasper H, Man J. *et al.* Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers: a systematic review and meta-analysis. *JAMA Oncol* 2018;**4**:522.
27. Hamada T, Kosumi K, Nakai Y. *et al.* Surrogate study endpoints in the era of cancer immunotherapy. *Ann Transl Med* 2018;**6**:S27. https://doi.org/10.21037/atm.2018.09.31.
28. Wang Y, Lai X, Wang J. *et al.* A joint model considering measurement errors for optimally identifying tumor mutation burden threshold. *Front Genet* 2022;**13**:915839. https://doi.org/10.3389/fgene.2022.915839.
29. Wang Y, Lai X, Wang J. *et al.* TMBcat: a multi-endpoint p-value criterion on different discrepancy metrics for superiorly inferring tumor mutation burden thresholds. *Front Immunol* 2022;**13**:995180. https://doi.org/10.3389/fimmu.2022.995180.
30. Yoshida Y, Kaneko M, Narukawa M. Magnitude of advantage in tumor response contributes to a better correlation between treatment effects on overall survival and progression-free survival: a literature-based meta-analysis of clinical trials in patients with metastatic colorectal cancer. *Int J Clin Oncol* 2020;**25**:851–60. https://doi.org/10.1007/s10147-020-01619-8.
31. Wang Y, Wang J, Fang W. *et al.* TMBserval: a statistical explainable learning model reveals weighted tumor mutation burden better categorizing therapeutic benefits. *Front Immunol* 2023;**14**:1151755. https://doi.org/10.3389/fimmu.2023.1151755.
32. Dondelinger F, Mukherjee S., The Alzheimer's Disease Neuroimaging Initiative. The joint lasso: high-dimensional regression for group structured data. *Biostatistics* 2020;**21**:219–35. https://doi.org/10.1093/biostatistics/kxy035.
33. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Series B Stat Methodology* 2006;**68**:49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x.
34. Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. *J R Stat Soc Series B Stat Methodol* 2014;**76**:373–97. https://doi.org/10.1111/rssb.12033.
35. Oates CJ, Korkola J, Gray JW. *et al.* Joint estimation of multiple related biological networks. *Ann Appl Stat* 2014;**8**:1892–919. https://doi.org/10.1214/14-AOAS761.
36. Tibshirani R, Saunders M, Rosset S. *et al.* Sparsity and smoothness via the fused lasso. *J R Stat Soc Series B Stat Methodol* 2005;**67**:91–108. https://doi.org/10.1111/j.1467-9868.2005.00490.x.
37. Ng SK, Tawiah R, Mclachlan GJ. *et al.* Joint frailty modeling of time-to-event data to elicit the evolution pathway of events: a generalized linear mixed model approach. *Biostatistics* 2022;**24**:108–23. https://doi.org/10.1093/biostatistics/kxab037.
38. Ye G-B, Xie X. Split Bregman method for large scale fused lasso. *Comput Stat Data Anal* 2011;**55**:1552–69. https://doi.org/10.1016/j.csda.2010.10.021.
39. Eisenhauer EA, Therasse P, Bogaerts J. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;**45**:228–47. https://doi.org/10.1016/j.ejca.2008.10.026.
40. Fang W, Yang Y, Ma Y. *et al.* Camrelizumab (SHR-1210) alone or in combination with gemcitabine plus cisplatin for nasopharyngeal carcinoma: results from two single-arm, phase 1 trials. *Lancet Oncol* 2018;**19**:1338–50. https://doi.org/10.1016/S1470-2045(18)30495-9.
41. Ma Y, Fang W, Zhang Y. *et al.* A phase I/II open-label study of nivolumab in previously treated advanced or recurrent nasopharyngeal carcinoma and other solid tumors. *Oncologist* 2019;**24**:891–e431. https://doi.org/10.1634/theoncologist.2019-0284.
42. Fang W, Ma Y, Yin JC. *et al.* Comprehensive genomic profiling identifies novel genetic predictors of response to anti–PD-(l)1 therapies in non–small cell lung cancer. *Clin Cancer Res* 2019;**25**:5015–26. https://doi.org/10.1158/1078-0432.CCR-19-0585.
43. Rizvi NA, Hellmann MD, Snyder A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* 2015;**348**:124–8. https://doi.org/10.1126/science.aaa1348.
44. Hellmann MD, Nathanson T, Rizvi H. *et al.* Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer.

cell carcinoma. *Science* 2018;**359**:801–6. https://doi.org/10.1126/science.aan5951.

52. Roth A, Khattra J, Yap D. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014;**11**: 396–8. https://doi.org/10.1038/nmeth.2883.

53. Chang T-G, Cao Y, Sfreddo HJ. *et al.* LORIS robustly predicts patient outcomes with immune checkpoint blockade therapy using common clinical, pathologic and genomic features. *Nat Cancer* 2024; **5**:1–18. https://doi.org/10.1038/s43018-024-00772-7.


*Briefings in Bioinformatics*, 2025, **26(1)**, bbae648
https://doi.org/10.1093/bib/bbae648
Problem Solving Protocol