Data-driven design of targeted gene panels for estimating immunotherapy biomarkers

Jacob R. Bradley, Timothy I. Cannings

January 2021

Abstract

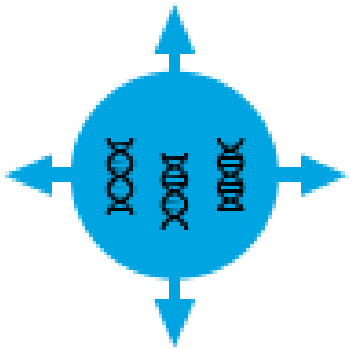
I. Problem: **Predict response to immunotherapy** with targeted gene panels.

II. Our approach: First fit a **generative model** of exome-wide mutation, then use this to construct efficient **predictive models** of biomarkers such as **Tumour Mutation Burden (TMB)**.

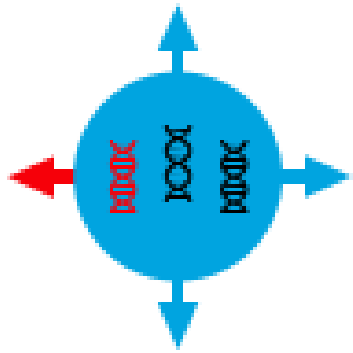
Outline

Cancer is a disease of the genome

Cancer arises when DNA in cells changes (mutates) [1].



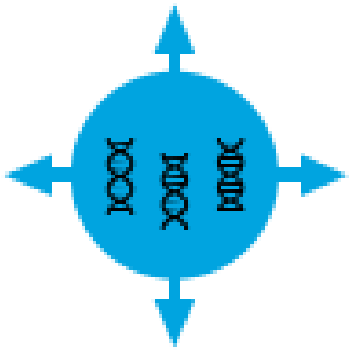
(a) Non-cancer cell: Normal DNA.



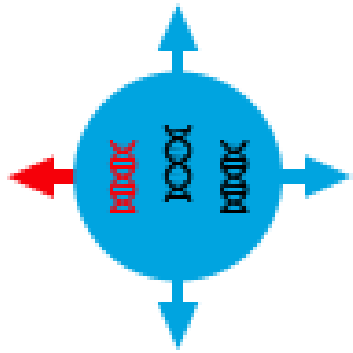
(b) Cancer cell: Mutated DNA.

Immunotherapy enables the immune system

Immunotherapy 'releases the brakes' on the immune system in order for it to attack tumours [2].



(a) Low-damage cell: **unrecognisable** to immune system.



(b) High-damage cell: **recognisable** to the immune system.

However, the immune system can only attack tumours that it recognises!

Tumour mutation burden stratifies patients for immunotherapy response

As a simple proxy for likelihood of immune response, we can use **tumour mutation burden**: the total count of non-synonymous mutations throughout the tumour exome [3, 4].

Advantages:

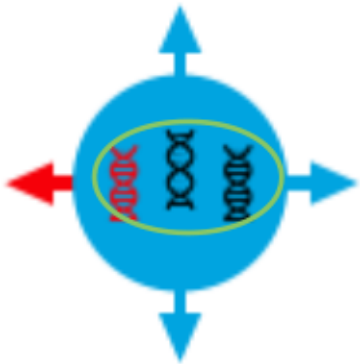
I. Calculated from DNA sequencing only (compatible with liquid biopsy [5]).
II. Relevant across cancer types.

Disadvantages:

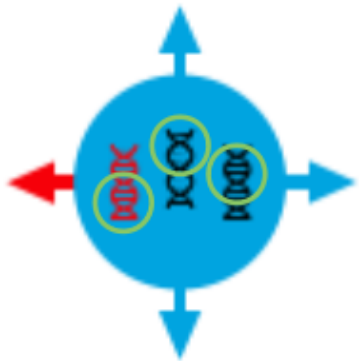
I. Requires Whole-Exome Sequencing (WES) to measure directly.
II. Fails to incorporate other features of the immune environment.

Targeted panels make genomic biomarkers viable

Rather than the entire exome/genome, targeted gene panels sequence a **subset** of genes.



(a) Whole-exome sequencing: **all genes** sequenced.



(b) Targeted panel sequencing: **subset** of genes sequenced.

Problem Statement

Aims:

1. Select a discrete set of genes to form a panel (of pre-specified cost).
2. Estimate TMB from panel sequencing data.

Mutations and biomarkers: some notation

We'll use i, g, and s to refer to:

1. A sample i (ranging from $i = 1$ to $i = N$),
2. A gene g (belonging to some index set G),
3. A variant type s (for example synonymous/non-synonymous).

Finally, we write:

4. M_{igs}, to refer to a mutation count of a particular sample, gene, and mutation type.

We then can define the TMB of the ith sample via:

$$T_i = \sum_{\substack{g \in G \\ s\text{: non-synonymous}}} M_{igs}.$$

Our generative model

A generative model of mutation attempts to capture the **underlying distribution** of variants across the entire exome/genome.

Our model incorporates **sample-specific** effects, **gene-specific** effects **variant-specific** effects and **variant-gene interactions**:

$$M_{igs} \sim \text{Poisson}(\phi_{igs}) \tag{1}$$

where

$$\log(\phi_{igs}) = \mu_i + \lambda_g + \eta_{gs} + \nu_s + \ell_g \tag{2}$$

NSCLC: Poisson-based generative model fit

We fit our generative model with a WES non-small cell lung cancer dataset from Campbell et al. (2016) [6].

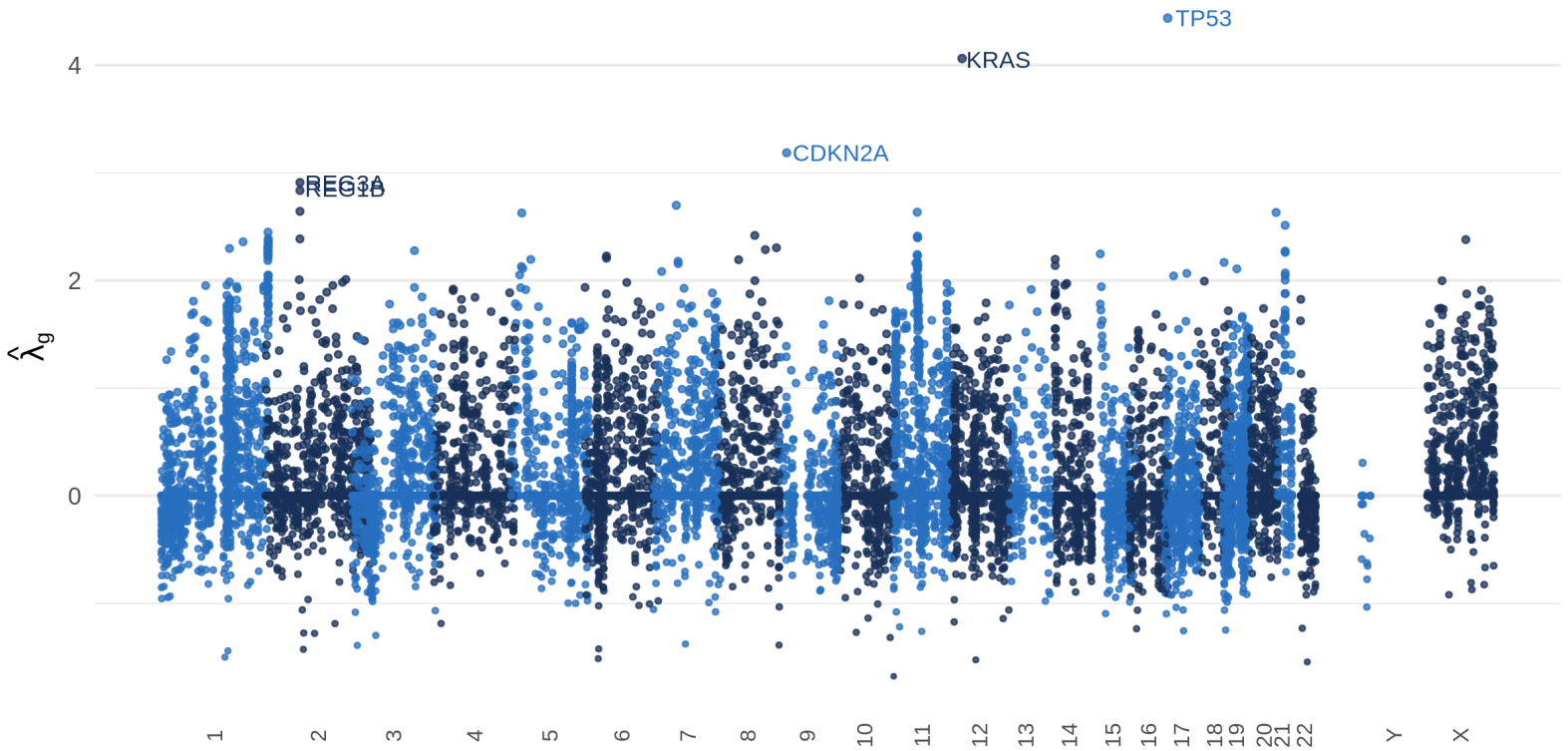


Figure: Generative model parameters. [7]

NSCLC: Poisson-based generative model fit

We fit our generative model with a WES non-small cell lung cancer dataset from Campbell et al. (2016) [6].

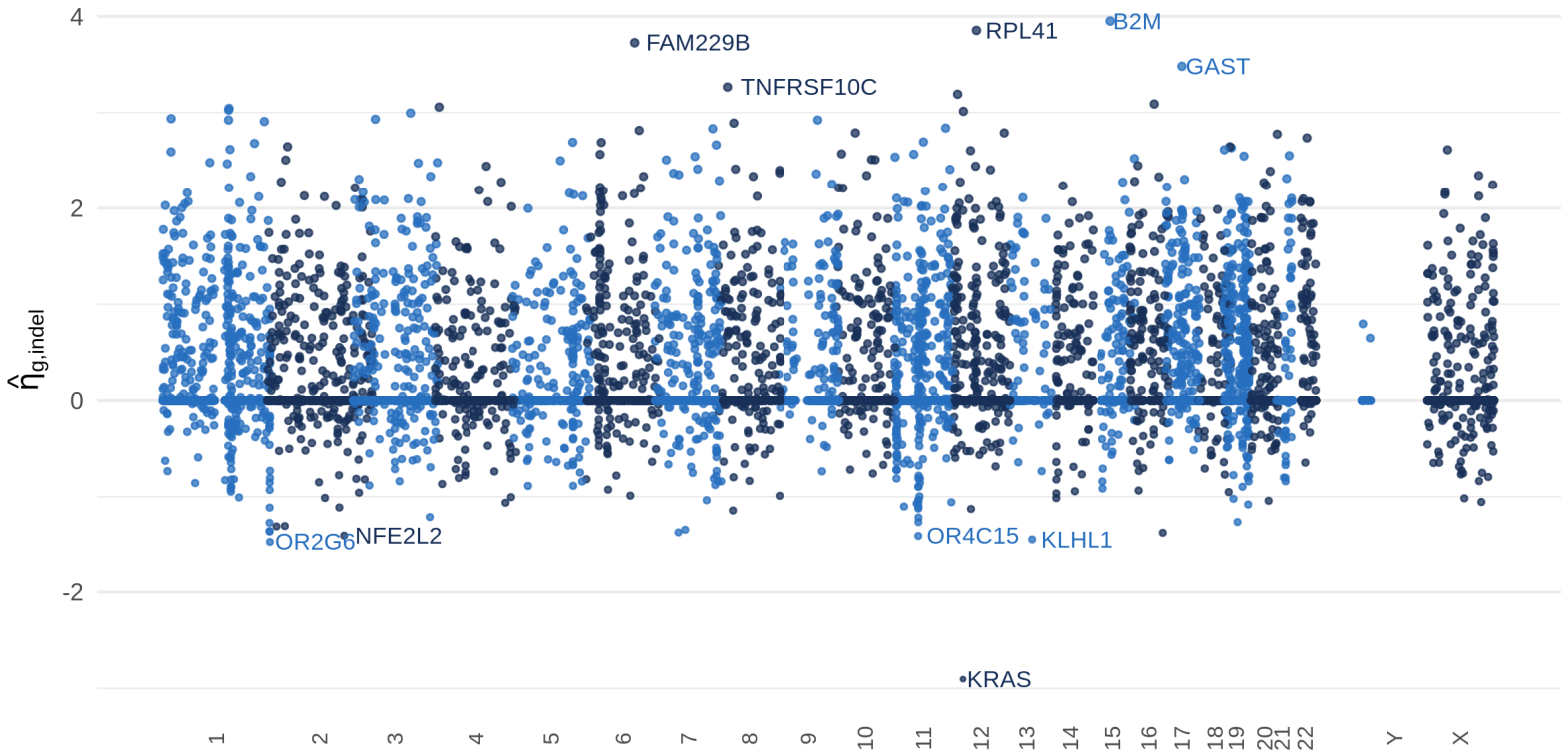


Figure: Generative model parameters (indel-specific) [7].

Proposed estimator

Aims:

1. Select a discrete set of genes to form a panel (of pre-specified cost).
2. Estimate TMB from panel sequencing data.

Our estimator:

$$\hat{T}_i(w) = \sum_{g,s} w_{gs} M_{igs} \tag{3}$$

How we fit our estimator:

$$\hat{w} := \arg\min_{w} \left\{ \mathbb{E}\left[(\hat{T}_i(w) - T_i)^2 \right] + \lambda |w| \right\}, \tag{4}$$

where λ is a penalty specifying panel cost, and the expectation $\mathbb{E}$ is over the distribution learned by our generative model.

NSCLC: Comparison panels and estimation methods

We compare our methodology with four commercial gene panels, and with three other means of predicting TMB. These panels are:

1. The TST-170 gene panel.
2. The Foundation One gene panel.
3. The MSK-IMPACT gene panel.
4. The TSO-500 gene panel.

and these methods are:

1. ecTMB ($\times$) [8].
2. Simple counting ($\circ$).
3. Linear regression ($+$).

NSCLC: Estimation performance

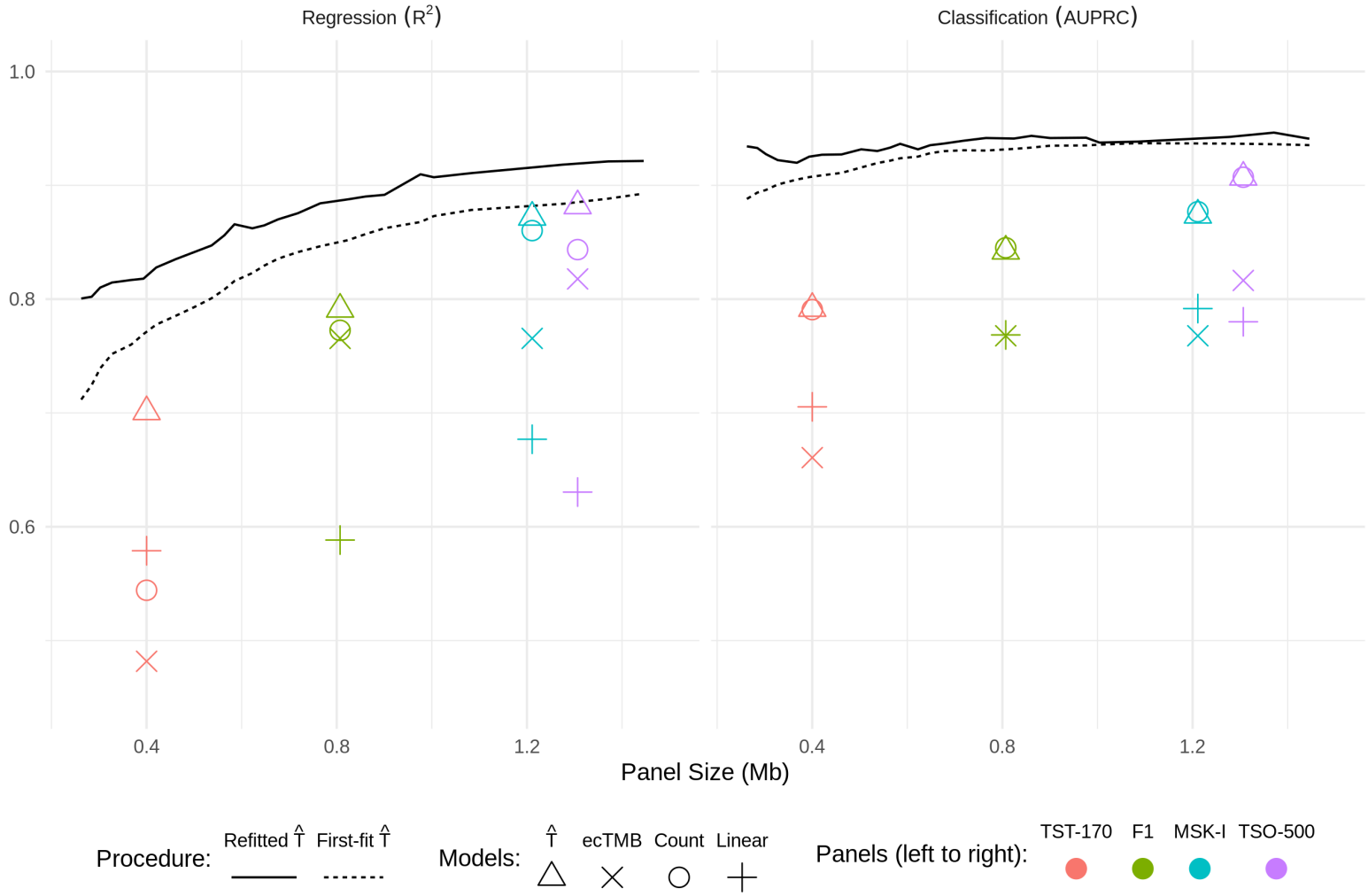


Figure: Predictive model performance: comparison with commercial panels and other predictive methods [7, 8].

Bonus material

Our method is flexible enough that we can:

1. Predict other exome-wide biomarkers, such as Tumour Indel Burden (TIB).
2. Augment existing gene panels to improve their predictive performance.

For more information check out our (draft) paper [7]!

References I

[1] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, March 2011.

[2] Drew M. Pardoll. The blockade of immune checkpoints in cancer immunotherapy. *Nature reviews. Cancer*, 12(4):252–264, March 2012.

[3] Jiaxin Zhu, Tiantian Zhang, Jiahao Li, Junming Lin, Wenhua Liang, Wenjie Huang, Ning Wan, and Jie Jiang. Association Between Tumor Mutation Burden (TMB) and Outcomes of Cancer Patients Treated With PD-1/PD-L1 Inhibitions: A Meta-Analysis. *Frontiers in Pharmacology*, 10, June 2019.

[4] Dedong Cao, Huilin Xu, Ximing Xu, Tao Guo, and Wei Ge. High tumor mutation burden predicts better efficacy of immunotherapy: a pooled analysis of 103078 cancer patients. *Oncoimmunology*, 8(9):e1629258, 2019.

References II

[5] David R. Gandara, Sarah M. Paul, Marcin Kowanetz, Erica Schleifman, Wei Zou, Yan Li, Achim Rittmeyer, Louis Fehrenbacher, Geoff Otto, Christine Malboeuf, Daniel S. Lieber, Doron Lipson, Jacob Silterra, Lukas Amler, Todd Riehl, Craig A. Cummings, Priti S. Hegde, Alan Sandler, Marcus Ballinger, David Fabrizio, Tony Mok, and David S. Shames. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nature Medicine*, 24(9):1441–1448, 2018.

[6] Joshua D. Campbell, Anton Alexandrov, Jaegil Kim, Jeremiah Wala, Alice H. Berger, Chandra Sekhar Pedamallu, Sachet A. Shukla, Guangwu Guo, Angela N. Brooks, Bradley A. Murray, Marcin Imielinski, Xin Hu, Shiyun Ling, Rehan Akbani, Mara Rosenberg, Carrie Cibulskis, Aruna Ramachandran, Eric A. Collisson, David J. Kwiatkowski, Michael S. Lawrence, John N. Weinstein, Roel G. W. Verhaak, Catherine J. Wu, Peter S. Hammerman, Andrew D. Cherniack, Gad Getz, Cancer Genome Atlas Research Network,

References III

Maxim N. Artyomov, Robert Schreiber, Ramaswamy Govindan, and Matthew Meyerson. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature Genetics*, 48(6):607–616, 2016.

[7] Jacob R. Bradley and Timothy I. Cannings. Data-driven design of targeted gene panels for estimating immunotherapy biomarkers, January 2021. https://github.com/cobrbra/TargetedPanelEstimation_Paper.

[8] Lijing Yao, Yao Fu, Marghoob Mohiyuddin, and Hugo Y. K. Lam. ecTMB: a robust method to estimate and classify tumor mutational burden. *Scientific Reports*, 10(1):1–10, March 2020. Number: 1 Publisher: Nature Publishing Group.