# Panels and Models for Accurate Prediction of Tumor Mutation Burden in Tumor Samples

Elizabeth Martínez Pérez[1], Miguel Angel Molina-Vila[2] and Cristina Marino-Buslje[1]
[1] Fundacion Instituto Leloir, Buenos Aires Argentina; [2] Laboratorio de Oncología/Pangaea Oncology, Quirón Dexeus. Barcelona, España.
Corresponding authors: emartinezperez1990@gmail.com; cmb@leloir.org.ar

## INTRODUCTION

In recent years, immunotherapies are a great promise as treatments for several cancers. An emerging biomarker for response to immunotherapy is the total number of nonsynonymous somatic mutations present in a tumor specimen, named tumor mutational burden (TMB) [1, 2]. Non synonymous mutations can generate neoantigens, which, in turn, can be recognized by the immune system, triggering an anticancer immune response [3].

High TMB correlates with a good response to checkpoint inhibitors. While treating cancer with immunotherapy can be highly effective, only some patients respond to this treatment, so, the challenge is to discriminate patients most likely to benefit from this therapy.

The goal of this work is to provide the subset of genes or exons needed to be sequenced in order to predict the TMB of cancer samples to make decisions about immunotherapy treatments and to reduce costs of sequencing. In such a way, to expand the number of people to which the evaluation could reach.

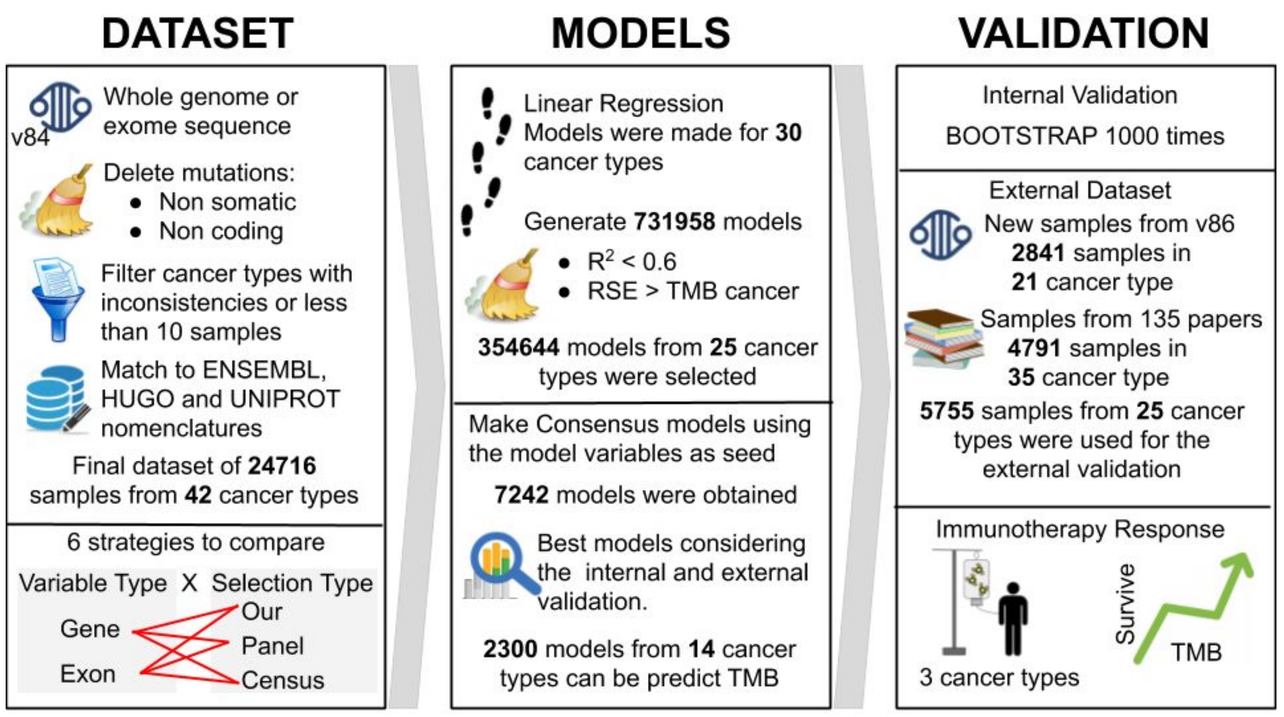


**Figure 1.** Methods Workflow

## RESULTS

### Mutational Burden in 42 cancer types.

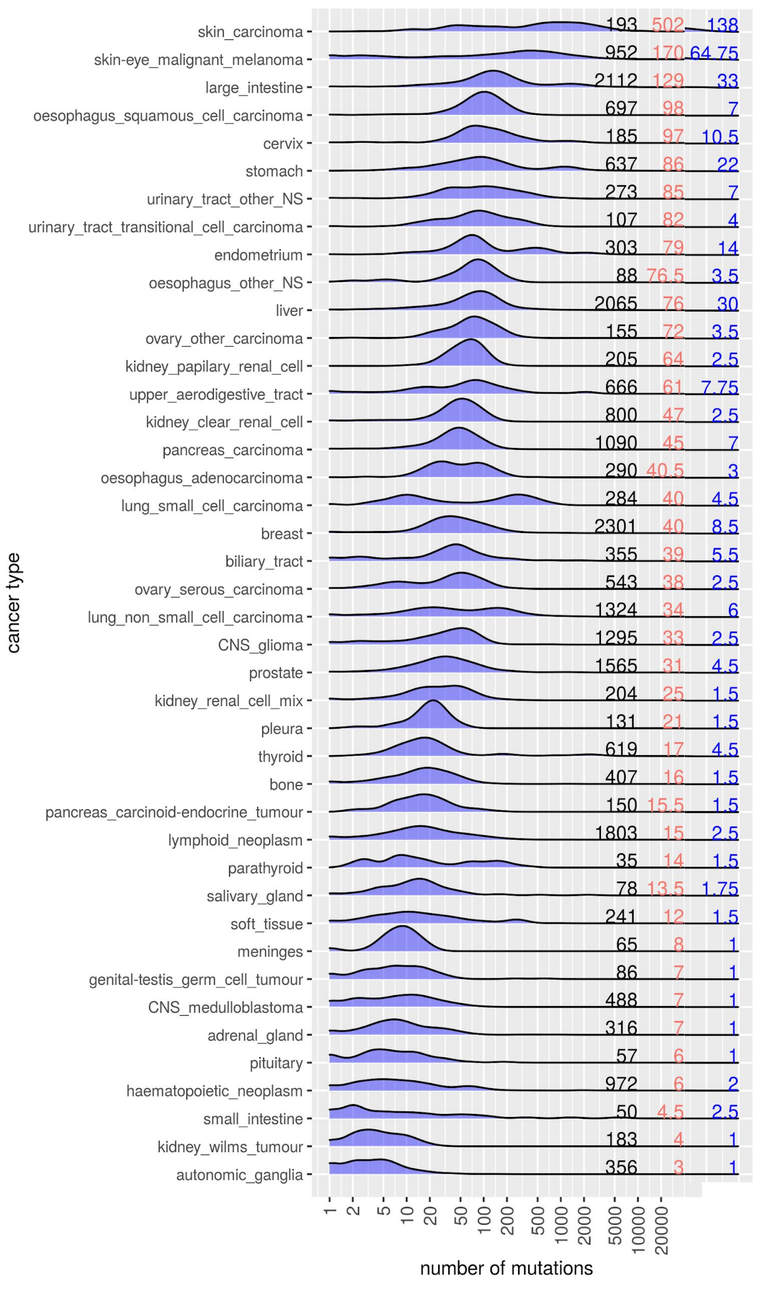


**Figure 2.** The "Y" axis for each cancer type is the frequency of mutated samples, the "X" axis is the TMB. Black numbers are the number of samples in the dataset. Red numbers are the median of TMB. Blue numbers are the median of genes in common. The cancer types are ordered by the median of mutational burden.

### TMB can be predicted with confidence in 14 cancer types

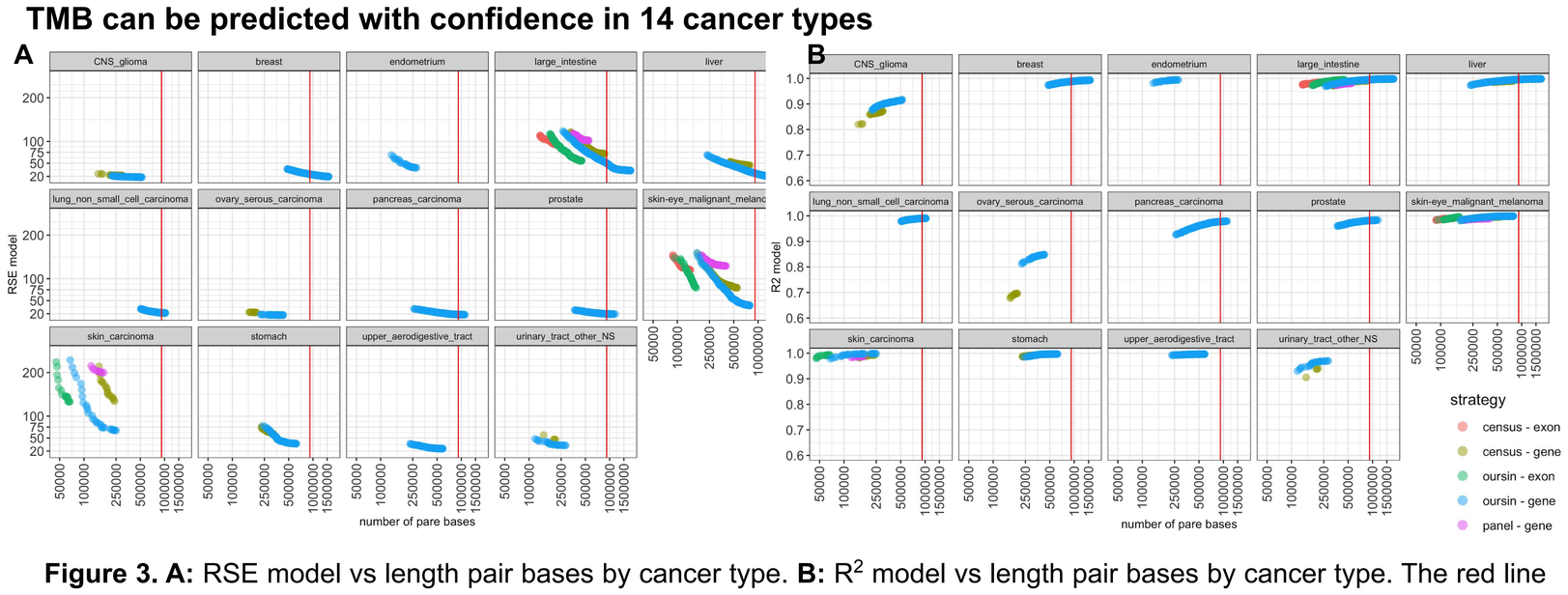


**Figure 3. A:** RSE model vs length pair bases by cancer type. **B:** $R^2$ model vs length pair bases by cancer type. The red line is the amount of pair bases in with the 315 genes panel

| Cancer Type | Final Mod | Number of genes | | | Length pair bases | | | RSE model | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Min | Max | Med | Min | Max | Med | Min | Max | Med |
| Breast | 161 | 48 | 210 | 130 | 479061 | 1562855 | 1045792 | 19. | 36 | 23 |
| CNS_glioma | 58 | 40 | 101 | 68 | 221314 | 524623 | 361376 | 17 | 21 | 18 |
| Endometrium | 25 | 16 | 40 | 28 | 136071 | 277732 | 211961 | 39 | 68 | 45 |
| Large intestine | 280 | 33 | 323 | 173 | 262885 | 1814292 | 1064564 | 33 | 123 | 40 |
| Liver | 250 | 26 | 278 | 150 | 233993 | 1737258 | 1031480 | 18 | 62 | 24 |
| Lung non small cell | 107 | 77 | 185 | 130 | 506478 | 1021231 | 785540 | 21 | 31 | 23 |
| Ovary serous | 30 | 33 | 67 | 50 | 225354 | 429223 | 328547 | 16 | 18 | 17 |
| Pancreas carcinoma | 130 | 30 | 170 | 94 | 259823 | 1109759 | 666645 | 17 | 31 | 19 |
| Prostate | 108 | 49 | 170 | 103 | 367953 | 1169163 | 756912 | 19 | 28 | 20 |
| Skin carcinoma | 31 | 9 | 40 | 24 | 67688 | 249940 | 142512 | 67 | 228 | 85 |
| Skin melanoma | 110 | 15 | 126 | 70 | 174743 | 787472 | 510966 | 38 | 159 | 52 |
| Stomach | 68 | 30 | 99 | 63 | 239251 | 621535 | 409077 | 37 | 77 | 42 |
| Upper aerodigestive tract | 61 | 36 | 96 | 66 | 231883 | 588199 | 431185 | 25 | 36 | 27 |
| Urinary tract other | 24 | 18 | 41 | 29 | 118624 | 284119 | 189464 | 32 | 47 | 35 |

**Table 1.** Statistics for models of the 14 cancer types with our-gene selection who pass all filters

### Genes emerging from model with our strategy

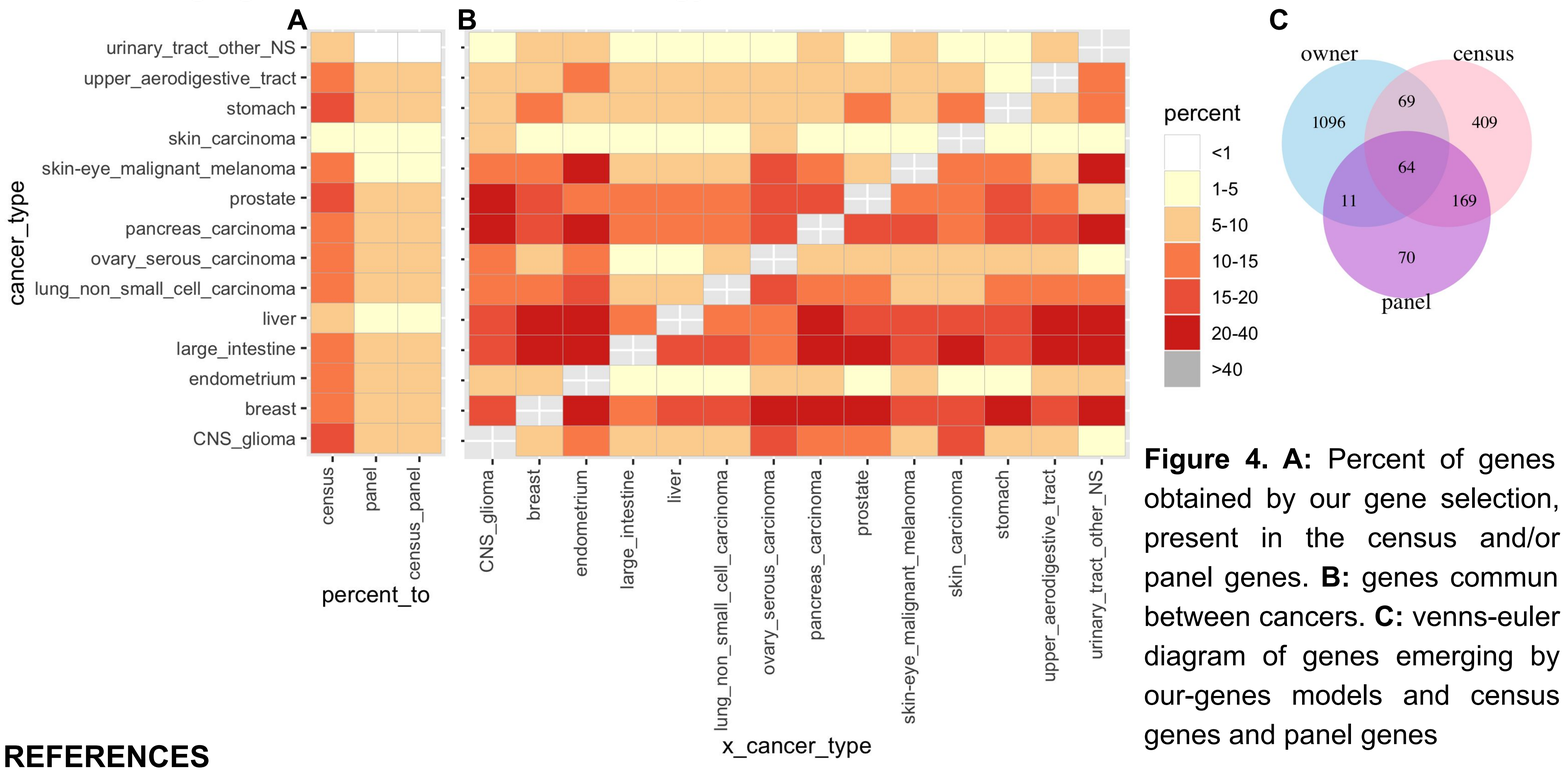


**Figure 4. A:** Percent of genes obtained by our gene selection, present in the census and/or panel genes. **B:** genes commun between cancers. **C:** venns-euler diagram of genes emerging by our-genes models and census genes and panel genes

## FUTURE WORK

- Correlation between models and response to immunotherapy is in process.
- Make a website with the obtained models per cancer types to be used for health professionals.

## CONCLUSIONS

- We can predict TMB with high accuracy in 14 cancer types, validated with internal and external datasets.
- Each cancer type has a different subset of genes from which the TMB can be predicted by knowing their mutations (see Figure 3B); questioning the usefulness of using a single panel for all the malignancies.
- Models obtained with our-gene selection perform better than the other selections and require a minor number of bases to be sequenced, reducing their cost.

## REFERENCES

1. Campbell BB et. al. Cell, 171:5, 987-989. 2. Chalmers ZR et. al. Genome Medicine. 2017. **9**:34 3. Colli LM et.al. Cancer Res. 2016.