BCB 731:

Defense Against the Dark Arts

Critic: A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy

November 1st, 2023

A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy

Marta Łuksza[1], Nadeem Riaz[2,3], Vladimir Makarov[3,4], Vinod P. Balachandran[5,6,7], Matthew D. Hellmann[7,8,9], Alexander Solovyov[10,11,12,13], Naiyer A. Rizvi[14], Taha Merghoub[7,15,16], Arnold J. Levine[1], Timothy A. Chan[2,3,4,7], Jedd D. Wolchok[7,8,15,16] & Benjamin D. Greenbaum[10,11,12,13]

Abstract

Checkpoint blockade immunotherapies enable the host immune system to recognize and destroy tumour cells[1]. Their clinical activity has been correlated with activated T-cell recognition of neoantigens, which are tumour-specific, mutated peptides presented on the surface of cancer cells[2,3]. Here we present a fitness model for tumours based on immune interactions of neoantigens that predicts response to immunotherapy. Two main factors determine neoantigen fitness: the likelihood of neoantigen presentation by the major histocompatibility complex (MHC) and subsequent recognition by T cells. We estimate these components using the relative MHC binding affinity of each neoantigen to its wild type and a nonlinear dependence on sequence similarity of neoantigens to known antigens. To describe the evolution of a heterogeneous tumour, we evaluate its fitness as a weighted effect of dominant neoantigens in the subclones of the tumour. Our model predicts survival in anti-CTLA-4-treated patients with melanoma[4,5] and anti-PD-1-treated patients with lung cancer[6]. Importantly, low-fitness neoantigens identified by our method may be leveraged for developing novel immunotherapies. By using an immune fitness model to study immunotherapy, we reveal broad similarities between the evolution of tumours and rapidly evolving pathogens[7,8,9].

Impact

Papers

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 10 May 2023

Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer

Luis A. Rojas, Zachary Sethna, Kevin C. Soares, Cristina Olcese, Nan Pang, Erin Patterson, Jayon Lihm, Nicholas Ceglia, Pablo Guasp, Alexander Chu, Rebecca Yu, Adrienne Kaya Chandra, Theresa Waters, Jennifer Ruan, Masataka Amisaki, Abderezak Zebboudj, Zagaa Odgerel, George Payne, Evelyna Derhovanessian, Felicitas Müller, Ina Rhee, Mahesh Yadav, Anton Dobrin, Michel Sadelain, ... Vinod P. Balachandran + Show authors

Nature **618**, 144–150 (2023) | Cite this article

195k Accesses | **66** Citations | **3620** Altmetric | Metrics

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 19 May 2022

Neoantigen quality predicts immunoediting in survivors of pancreatic cancer

Marta Łuksza, Zachary M. Sethna, Luis A. Rojas, Jayon Lihm, Barbara Bravi, Yuval Elhanati, Kevin Soares, Masataka Amisaki, Anton Dobrin, David Hoyos, Pablo Guasp, Abderezak Zebboudj, Rebecca Yu, Adrienne Kaya Chandra, Theresa Waters, Zagaa Odgerel, Joanne Leung, Rajya Kappagantula, Alvin Makohon-Moore, Amber Johns, Anthony Gill, Mathieu Gigoux, Jedd Wolchok, Taha Merghoub, ... Vinod P. Balachandran + Show authors

Nature **606**, 389–395 (2022) | Cite this article

42k Accesses | **53** Citations | **219** Altmetric | Metrics

nature

Explore content | About the journal | Publish with us

nature > articles > article

Article | Open access | Published: 11 May 2022

Fundamental immune–oncogenicity trade-offs define driver mutation fitness

David Hoyos, Roberta Zappasodi, Isabell Schulze, Zachary Sethna, Kelvin César de Andrade, Dean F. Bajorin, Chaitanya Bandlamudi, Margaret K. Callahan, Samuel A. Funt, Sine R. Hadrup, Jeppe S. Holm, Jonathan E. Rosenberg, Sohrab P. Shah, Ignacio Vázquez-García, Britta Weigelt, Michelle Wu, Dmitriy Zamarin, Laura F. Campitelli, Edward J. Osborne, Mark Klinger, Harlan S. Robins, Payal P. Khincha, Sharon A. Savage, Vinod P. Balachandran, ... Benjamin D. Greenbaum + Show authors

Published: 01 November 2017

Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer

Vinod P. Balachandran, Marta Łuksza, Julia N. Zhao, Vladimir Makarov, John Alec Moral, Romain Remark, Brian Herbst, Gokce Askan, Umesh Bhanot, Yasin Senbabaoglu, Daniel K. Wells, Charles Ian Ormsby Cary, Olivera Grbovic-Huezo, Marc Attiyeh, Benjamin Medina, Jennifer Zhang, Jennifer Loo, Joseph Saglimbeni, Mohsen Abu-Akeel, Roberta Zappasodi, Nadeem Riaz, Martin Smoragiewicz, Z. Larkin Kelley, Olca Basturk, Australian Pancreatic Cancer Genome Initiative, ... Steven D. Leach + Show authors

Business & Money

STAND UP TO CANCER ANNOUNCES $1.5 MILLION COMMITMENT FROM PANCREATIC CANCER NORTH AMERICA TO FUND PANCREATIC CANCER VACCINE RESEARCH

Funding

STAND UP TO CANCER

Pershing Square Sohn Prize-Mark Foundation Fellow Announced

COMPUTATIONAL BIOLOGIST AWARDED PRESTIGIOUS RESEARCH PRIZE

In partnership with the Pershing Square Sohn Cancer Research Alliance, The Mark Foundation is pleased to announce that **Dr. Benjamin Greenbaum** has been named the 2018 Pershing Square Sohn Prize-Mark Foundation Fellow. This award will support Dr. Greenbaum's research, in which he uses methods from math and physics to analyze highly complex data to understand better immunotherapy responses so that more patients can be successfully treated.

Plausibility of the Neoantigen Fitness Model

The Model

$$n(\tau) = \sum_{\alpha} X_{\alpha} \exp\left(-\max_{i \in \text{ clone } \alpha} (A_i \times R_i)\, \tau\right) \quad (5)$$

Relative Tumor Size In The Future

α: tumor sub-clone

X_α: Initial clonal fraction

i: Mutant peptide with high predicted MHC affinity

A_i: Will peptide outcompete other peptides to be presented on tumor cell MHCs?

R_i: Will peptide-MHC complex be recognized by T-cells?

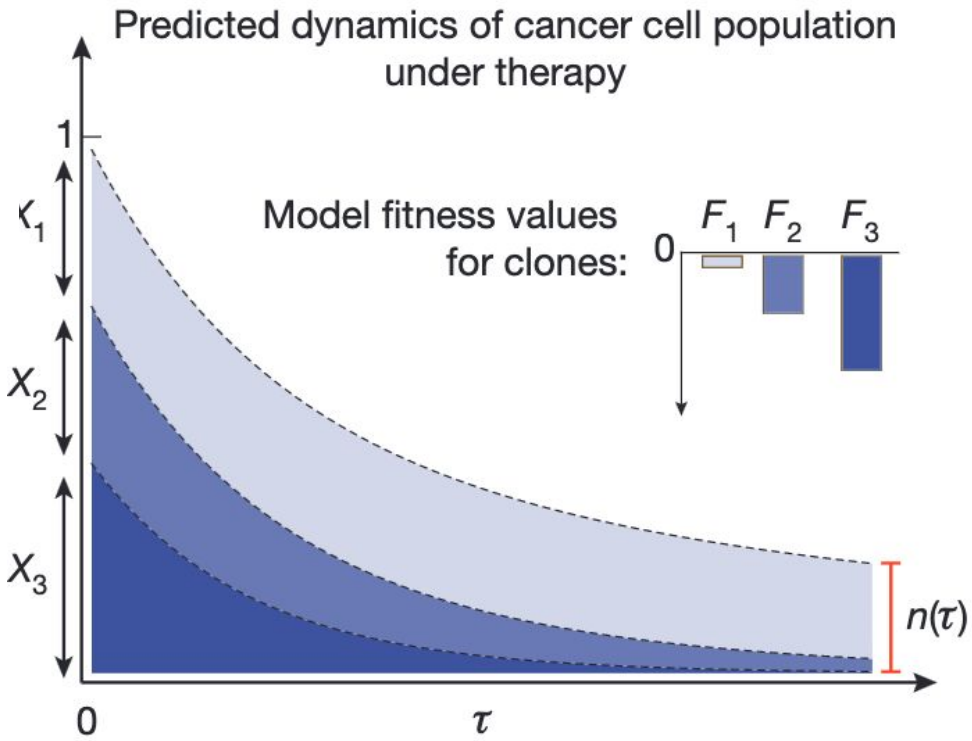
$$A = K_{\mathrm{d}}^{\mathrm{WT}} / K_{\mathrm{d}}^{\mathrm{MT}}$$

$$R = Z(k)^{-1} \sum_{e \in \mathrm{IEDB}} \exp(-k(a - |\boldsymbol{s}, \boldsymbol{e}|))$$

Overall form: tumors always shrink

$$n(\tau) = \sum_{\alpha} X_{\alpha} \exp(F_{\alpha}\tau)$$

Always negative



“Amplitude” / “Agretopicity” / “DAI”

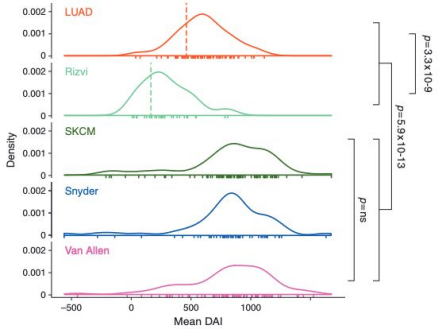
$$A = K_d^{WT} / K_d^{MT} \quad (7)$$

Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity

Fei Duan,[1] Jorge Duitama,[2] Sahar Al Seesi,[2] Cory M. Ayres,[3] Steven A. Corcelli,[3] Arpita P. Pawashe,[1] Tatiana Blanchard,[1] David McMahon,[1] John Sidney,[4] Alessandro Sette,[4] Brian M. Baker,[3] Ion I. Mandoiu,[2] and Pramod K. Srivastava[1]

Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma

E. Ghorani[1,2], R. Rosenthal[2], N. McGranahan[2,3], J. L. Reading[1,2], M. Lynch[4], K. S. Peggs[1,2], C. Swanton[2,3*,†] & S. A. Quezada[1,2*,†]



Which features matter?

Resource

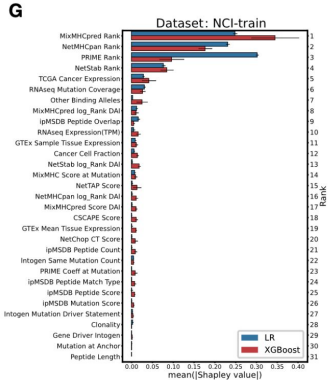
Machine learning methods and harmonized datasets improve immunogenic neoantigen prediction

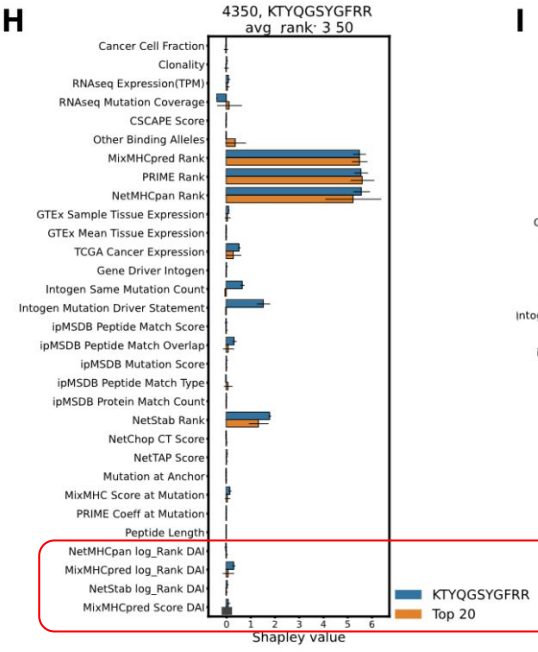
Markus Müller,[1,2,3,4,*] Florian Huber,[1,2,3] Marion Arnaud,[1,2,3] Anne I. Kraemer,[1,2,3] Emma Ricart Altimiras,[1,2,3] Justine Michaux,[1,2,3] Marie Taillandier-Coindard,[1,2,3] Johanna Chiffelle,[1,2,3] Baptiste Murgues,[1,2,3] Talita Gehret,[1,2,3] Aymeric Auger,[1,2,3] Brian J. Stevenson,[5,*] George Coukos,[1,2,3,6] Alexandre Harari,[1,2,3,6] and Michal Bassani-Sternberg[1,2,3,3,4,*]

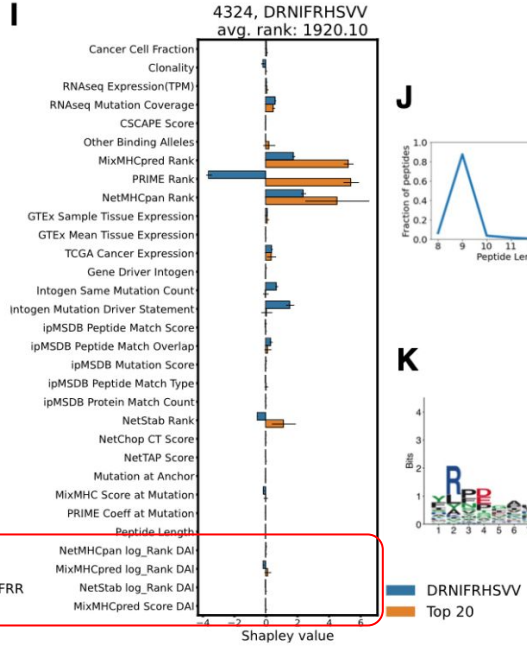
[1]Ludwig Institute for Cancer Research, University of Lausanne, Agora Center Bugnon 25A, 1005 Lausanne, Switzerland
[2]Department of Oncology, Centre hospitalier universitaire vaudois (CHUV), Rue du Bugnon 46, 1005 Lausanne, Switzerland
[3]Agora Cancer Research Centre, 1011 Lausanne, Switzerland
[4]SIB Swiss Institute of Bioinformatics, Quartier Sorge, Bâtiment Amphipôle, 1015 Lausanne, Switzerland
[5]Center of Experimental Therapeutics, Department of Oncology, Centre hospitalier universitaire vaudois (CHUV), Rue du Bugnon 46, 1005 Lausanne, Switzerland
[6]Lead contact
*Correspondence: markus.muller@chuv.ch (M.M.), michal.bassani@chuv.ch (M.B.-S.)
https://doi.org/10.1016/j.immuni.2023.09.002

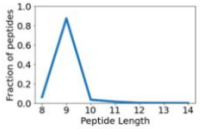
SUMMARY

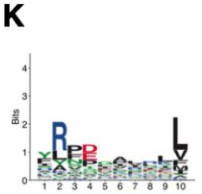
The accurate selection of neoantigens that bind to class I human leukocyte antigen (HLA) and are recognized by autologous T cells is a crucial step in many cancer immunotherapy pipelines. We reprocessed whole-exome sequencing and RNA sequencing (RNA-seq) data from 120 cancer patients from two external large-scale neoantigen immunogenicity screening assays combined with an in-house dataset of 11 patients and identified 46,017 somatic single-nucleotide variant mutations and 1,781,445 neo-peptides, of which 212 mutations and 178 neo-peptides were immunogenic. Beyond features commonly used for neoantigen prioritization, factors such as the location of neo-peptides within protein HLA presentation hotspots, binding promiscuity, and the role of the mutated gene in oncogenicity were predictive for immunogenicity. The classifiers accurately predicted neoantigen immunogenicity across datasets and improved their ranking by up to 30%. Besides insights into machine learning methods for neoantigen ranking, we have provided homogenized datasets valuable for developing and benchmarking companion algorithms for neoantigen-based immunotherapies.



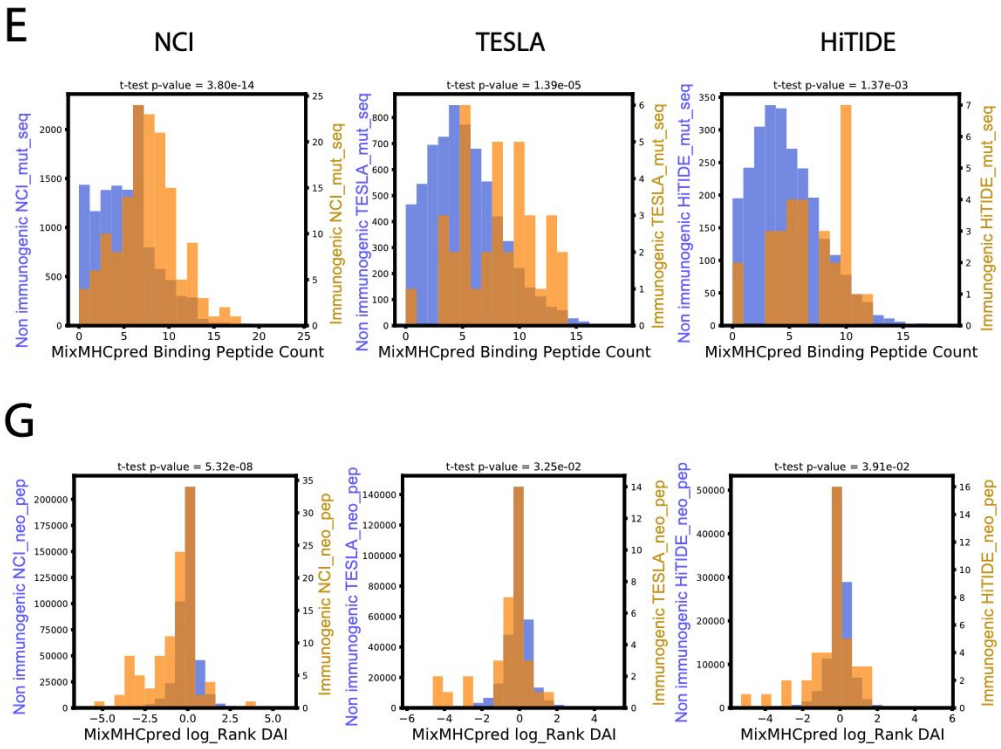




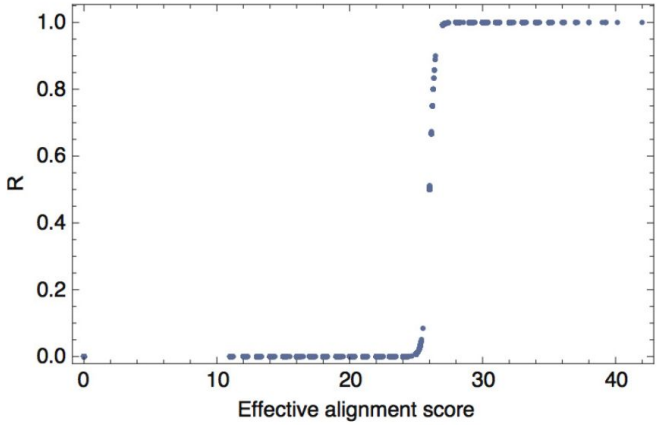




“Amplitude” / “Agretopicity” / “DAI”



T-cell recognition potential



Extended Data Figure 4 | Alignments to IEDB epitopes. The TCR recognition probability for a neoantigen is a sigmoidal function of the alignment scores of a given neoantigen to the IEDB epitopes, evaluated for the set of neoantigens from the cohort of patients from ref. 5, using the consistent set of parameters.

$$R = Z(k)^{-1} \sum_{e \in \mathrm{IEDB}} \exp(-k(a - |s, e|))$$

Pathogen derived T-cell epitopes

Alignment score cutoff

Weighted edit distance using BLOSUM62

What is IEDB?

IEDB ID	Reference	Epitope	Host	Immunization	Assay Antigen	Antigen Epitope Relation	MHC Restriction	Assay Description
1333767	Susanne Tartz; Infect Immun 2006	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	No immunization was performed	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**
1384778	Brenna Carey; Clin Immunol 2005	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	No immunization was performed	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**
1813240	M Baird; Scand J Immunol 2004	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	No immunization was performed	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**
1818239	Maiko Taneichi; PLoS One 2010	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus C57BL/6 X BALB/c	Administration in vivo with SIINFEKL (Epitope)	ovalbumin ovalbumin Gallus gallus (chicken)	Source Antigen	H2-Kb	3H-thymidine proliferation **Positive**
1954895	Lesley A Smyth; J Immunol 2012	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	No immunization was performed	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**
1980620	Yongliang Zhang; J Biol Chem 2009	SIINFEKL Ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus (mouse)	Administration in vivo with Ovalbumin (Source Antigen)	SIINFEKL Ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2 class I	3H-thymidine proliferation **Positive**
1884022	Bruno Garulli; J Biomed Biotechnol 2011	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	The immunization procedure is unknown	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-b class I	3H-thymidine proliferation **Positive**
2117814	Maria del Rosario Espinoza Mora; PLoS One 2014	SIINFEKL Ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus OT-1 TCR Tg	No immunization was performed	SIINFEKL Ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**
1384775	Brenna Carey; Clin Immunol 2005	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Mus musculus C57BL/6	Administration in vivo with SIINFEKL (Epitope)	SIINFEKL ovalbumin (258-265) Gallus gallus (chicken)	Epitope	H2-Kb	3H-thymidine proliferation **Positive**

How big is IEDB?

Summary Metrics	
Peptidic Epitopes	1,598,256
Non-Peptidic Epitopes	3,188
T Cell Assays	509,612
B Cell Assays	1,391,866
MHC Ligand Assays	4,778,192
Epitope Source Organisms	4,405
Restricting MHC Alleles	995
References	24,286

Paper only uses 6,695 IEDB entries

```
(base) ➜  SupplementaryDataFile5 cat IEDB_negative_T-cell_assays.fasta | grep ">" | wc -l
    4137
(base) ➜  SupplementaryDataFile5 cat IEDB_positive_T-cell_assays.fasta | grep ">" | wc -l
    2558
```

Which IEDB entries do they use?

```
(base) → SupplementaryDataFile5 cat IEDB*.fasta | grep corona
>16351|Spike glycoprotein precursor|SARS coronavirus|227859
>18160|Spike glycoprotein precursor|SARS coronavirus BJ01|228407
>28784|Spike glycoprotein precursor|Human coronavirus 229E|11137
>32036|Spike glycoprotein precursor|SARS coronavirus BJ01|228407
>37515|Nucleoprotein|SARS coronavirus BJ01|228407
>54725|Spike glycoprotein precursor|SARS coronavirus BJ01|228407
>74683|Nucleoprotein|SARS coronavirus BJ01|228407
>16156|Spike glycoprotein precursor|P59594.1|SARS coronavirus|227859
>19442|nucleocapsid protein|ABI96968.1|SARS coronavirus|227859
>21041|Membrane glycoprotein|Q692E0|SARS coronavirus TJF|284672
>21347|Nucleoprotein|P59595.1|SARS coronavirus|227859
>26273|nucleocapsid protein|ABI96968.1|SARS coronavirus|227859
>27241|Spike glycoprotein precursor|P59594.1|SARS coronavirus|227859
>32069|Spike glycoprotein precursor|P59594.1|SARS coronavirus BJ01|228407
>36723|Spike glycoprotein precursor|P15423.1|Human coronavirus 229E|11137
>36724|Spike glycoprotein precursor|P59594.1|SARS coronavirus|227859
>37473|Nucleoprotein|P59595.1|SARS coronavirus|227859
>37536|Spike glycoprotein precursor|P15423.1|Human coronavirus 229E|11137
>50779|N protein|AAP13445.1|SARS coronavirus Urbani|228330
>51250|nucleocapsid protein|ABI96968.1|SARS coronavirus|227859
>54680|Spike glycoprotein precursor|P59594.1|SARS coronavirus|227859
>54690|Nucleoprotein|P59595.1|SARS coronavirus|227859
>55683|nucleocapsid protein|ABI96968.1|SARS coronavirus|227859
>64710|Membrane glycoprotein|Q692E0|SARS coronavirus TJF|284672
>71663|Spike glycoprotein precursor|P59594.1|SARS coronavirus|227859
>112359|Non-structural protein 2a|Q80872.1|Human coronavirus OC43|31631
>156949|Protein 3a|P59632.1|SARS coronavirus|227859
>190494|nucleocapsid protein|NP_828858.1|SARS coronavirus|227859
>190533|nucleocapsid protein|AAP49024.1|SARS coronavirus|227859
>193551|Protein 3a|P59632.1|SARS coronavirus|227859
>532115|membrane glycoprotein|ADE34769.1|SARS coronavirus|227859
>534068|nucleocapsid protein|NP_828858.1|SARS coronavirus|227859
```

```
 93 Trypanosoma cruzi
  1 Trypanosoma cruzi strain CL Brener
 39 West Nile virus
 12 West Nile virus NY-99
```

```
 11 Dengue virus
 79 Dengue virus 1
  2 Dengue virus 1 Singapore/S275/1990
158 Dengue virus 2
 33 Dengue virus 2 D2/SG/05K4155DK1/2005
 13 Dengue virus 2 Thailand/16681/84
  4 Dengue virus 2 Thailand/NGS-C/1944
 46 Dengue virus 3
 43 Dengue virus 4
  1 Dengue virus 4 Thailand/0348/1991
 21 Dengue virus type 1 Hawaii
```

```
14 Japanese encephalitis virus
 1 Japanese encephalitis virus Vellore P20778
```

Red Flags

CNV & phylogeny from WES

Zhao *et al. BMC Bioinformatics* (2020) 21:97
https://doi.org/10.1186/s12859-020-3421-1

BMC Bioinformatics

RESEARCH ARTICLE Open Access

Comparative study of whole exome sequencing-based copy number variation detection tools

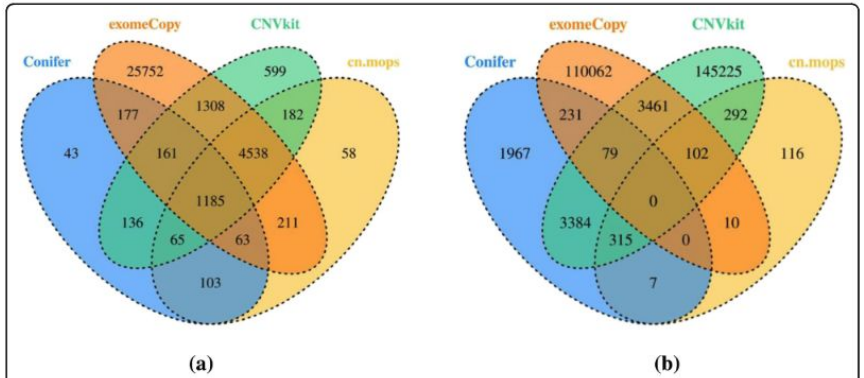


Fig. 4 The overlapping consistency results. Fig **a** and **b** show the simulated data and real data.

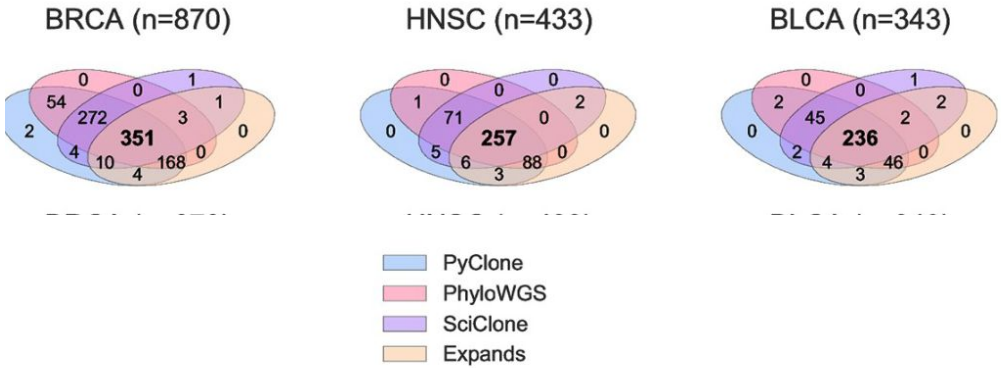
RESEARCH ARTICLE

Assessing reliability of intra-tumor heterogeneity estimates from single sample whole exome sequencing data

Judith Abécassis[1,2,3], Anne-Sophie Hamy[1], Cécile Laurent[1], Benjamin Sadacca[1,4], Hélène Bonsang-Kitzis[1,5], Fabien Reyal[1,5], Jean-Philippe Vert[2,6]*

1 Institut Curie, PSL Research University, Translational Research Department, INSERM, U932 Immunity and Cancer, Residual Tumor & Response to Treatment Laboratory (RT2Lab), Paris, France, **2** MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France, **3** Institut Curie, PSL Research University, INSERM, U900, Paris, France, **4** Institut de Mathématiques de Toulouse, UMR5219 Université de Toulouse, CNRS UPS IMT, Toulouse, France, **5** Department of Surgery, Institut Curie, Paris, France, **6** Google Brain, Paris, France

* jpvert@google.com



Where's the code, Lebowski?

Code availability. Custom script examples for computation of neoantigen fitness cost are included as Supplementary Data 7. Additional custom code will be made available upon reasonable request.

```
##############################################################################################
# Compute neoantigen fitness cost (recognition potential)
#
# Directory structure:
#
# InputData:
#   1. neoantigen-data files for 3 cohorts (Van Allen et al., Snyder et al. and Rizvi et al.)
#   2. iedb.fasta: fasta file with IEDB epitope sequences
#   3. neoantigen alignment folders with precomputed blastp alignments for all neoantigens, split into files for each sample.
#    eg. run as
#    blastp -query <InputData/neoantigen_alignments_Rizvi/neoantigens_AL4602.fasta> -db InputData/iedb.fasta -outfmt 5 -evalue 100000000  -gapopen 11 -gapextend 1 > <InputData/neoantigen_alignments_Rizvi/neoantigens_AL4602.xml>
#
# src:
#   source code folder
#
# Output:
#   source code output folder: neoantigens with computed fitness cost
##############################################################################################
#
# fitness model parameters
a=26.
k=4.86936

python src/main.py InputData/neoantigens_VanAllen.txt InputData/neoantigen_alignments_VanAllen $a $k Output/neoantigen_fitness_VanAllen.txt
python src/main.py InputData/neoantigens_Snyder.txt InputData/neoantigen_alignments_Snyder $a $k Output/neoantigen_fitness_Snyder.txt
python src/main.py InputData/neoantigens_Rizvi.txt InputData/neoantigen_alignments_Rizvi $a $k Output/neoantigen_fitness_Rizvi.txt
```

They evaluate *lots* of models

Extended Data Table 2 | Ranking of fitness models without accounting for the subclonal composition c

Melanoma, anti-CTLA-4 (ref .5)	Parameters trained on melanoma anti-CTLA-4 dataset from ref. 4				Log-rank test			
Models	*a*		*k*		Score		Significance	Equation
	Mean	Std	Mean	Std	Mean	Std		
$A \times R$	18.5	±0.152	0.59845	±0.001	0.61	±0.24		(2)
Partial models:								
A	-	-	-	-	0.05	±0.03		(13)
R	18.2	±2.995	4.54981	±0.001	0.2	±0.03		(14)
$K_d^{WT} \times R$	23.7	±2.593	3.83397	±0.001	1.35	±0.34		(15)
K_d^{WT}	-	-	-	-	0.46	±0.2		(15)
$1/K_d^{MT} \times R$	26.6	±1.609	1.34421	±0.001	1.41	±0.61		(16)
$1/K_d^{MT}$	-	-	-	-	0.69	±0.23		(16)
Alternative models:								
Neoantigen load	-	-	-	-	0.24	±0.12		(18)
$A \times R$, sum over neoantigens	25.1	±3.308	4.89176	±0.001	1.57	±0.39		(17)
$A \times R$, alignments at positions 3-8	21.9	±2.871	3.1276	±0.001	0.35	±0.36		(2)
$A \times R$, negative IEDB assays	14.4	±2.58	2.20223	±0.001	0.04	±0.03		(2)
$A \times R$, all neoantigens	29.2	±2.892	3.88642	±0.001	0.24	±0.19		(2)

Melanoma, anti-CTLA-4 (ref. 4)	Parameters trained on melanoma anti-CTLA-4 dataset from ref. 5				Log-rank test			
Models	*a*		*k*		Score		Significance	Equation
	Mean	Std	Mean	Std	Mean	Std		
$A \times R$	26.4	±0.892	1.0851	±0.001	6.55	±0.9	0.01047 *	(2)
Partial models:								
A	-	-	-	-	4.44	±0.68	0.03507 *	(13)
R	29.7	±4.829	2.51962	±0.001	1.26	±0.34		(14)
$K_d^{WT} \times R$	26.9	±2.735	3.73387	±0.001	1.67	±0.22		(15)
K_d^{WT}	-	-	-	-	3.11	±0.75		(15)
$1/K_d^{MT} \times R$	26	±1.929	0.82074	±0.001	3.65	±0.81		(16)
$1/K_d^{MT}$	-	-	-	-	3.44	±0.61		(16)
Alternative models:								
Neoantigen load	-	-	-	-	0.42	±0.21		(18)
$A \times R$, sum over neoantigens	27	±3.005	5.08369	±0.001	1.63	±0.53		(17)
$A \times R$, alignments at positions 3-8	30	±4.023	1.35553	±0.001	2.73	±0.65		(2)
$A \times R$, negative IEDB assays	36	±9.57	10.1531	±0.001	0.23	±0.65		(2)
$A \times R$, all neoantigens	26	±1.95	5.22216	±0.001	0.59	±0.92		(2)

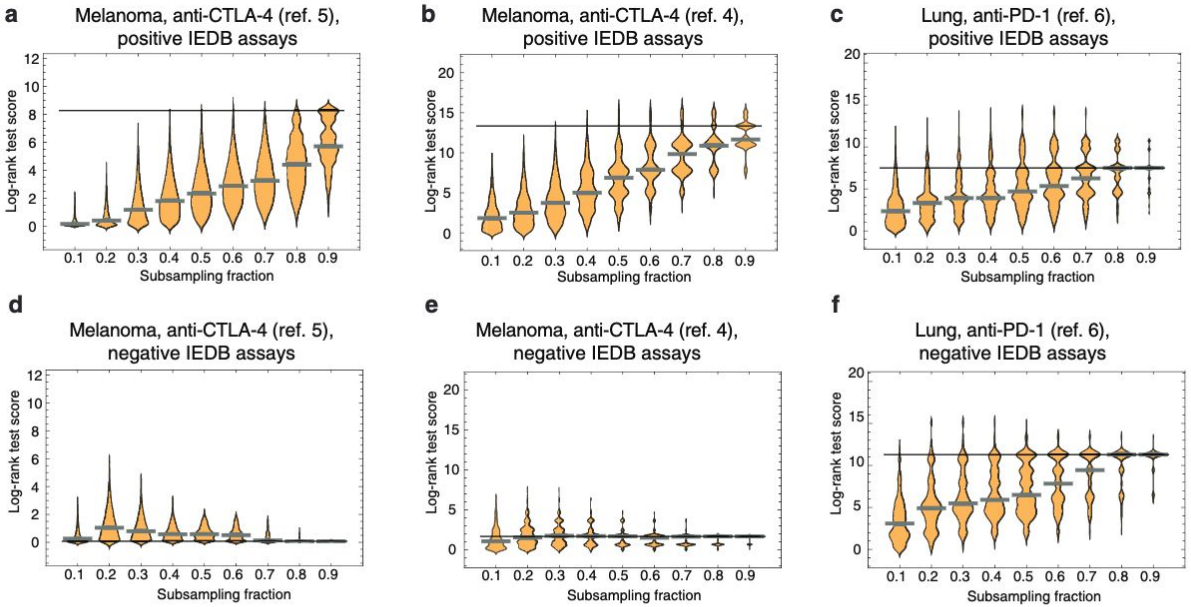
Lung, anti-PD-1 (ref. 6)	Parameters trained on melanoma anti-CTLA-4 datasets from ref. 4 and ref. 5				Log-rank test			
Models	*a*		*k*		Score		Significance	Equation
	Mean	Std	Mean	Std	Mean	Std		
$A \times R$	27	±0.787	1.00032	±0.001	6.48	±1.14	0.01088 *	(2)
Partial models:								
A	-	-	-	-	4.65	±1.17	0.03099 *	(13)
R	19.6	±3.355	4.29127	±0.001	1.53	±0.29		(14)
$K_d^{WT} \times R$	23	±2.737	5.37707	±0.001	10.48	±1.71	0.00121 ***	(15)
K_d^{WT}	-	-	-	-	4.49	±0.75	0.03416 *	(15)
$1/K_d^{MT} \times R$	21	±2.027	0.53498	±0.001	0.02	±0.07		(16)
$1/K_d^{MT}$	-	-	-	-	0.17	±0.13		(16)
Alternative models:								
Neoantigen load	-	-	-	-	4.93	±1.15	0.02639 *	(18)
$A \times R$, sum over neoantigens	25	±4.316	4.03113	±0.001	3.09	±1.09		(17)
$A \times R$, alignments at positions 3-8	22	±3.042	5.2228	±0.001	2.3	±0.82		(2)
$A \times R$, negative IEDB assays	14.5	±2.806	1.90577	±0.001	1	±0.45		(2)
$A \times R$, all neoantigens	29	±4.469	1.87092	±0.001	0.75	±0.62		(2)

Melanoma, anti-CTLA-4 (ref .5)	Parameters trained on melanoma, anti-CTLA-4 dataset from ref. 4						Log-rank test			
Models	τ		*a*		*k*		Score		Significance	Equation
	Mean	Std	Mean	Std	Mean	Std	Mean	Std		
$A \times R$	0.09003	±0.077	26	±2.988	4.19761	±0.01	7.92	±0.49	0.00489 ***	(2)
Partial models:										
A	0.00131	±0.0001	-	-	-	-	0.65	±0.03		(13)
R	12.33338	±1.827	12.5	±1.074	1.89795	±24.161	1.68	±0.01		(14)
$K_d^{WT} \times R$	0.00048	±0.0001	31	±2.911	6.26907	±0.001	0.09	±0.25		(15)
K_d^{WT}	0.00307	±0.0001	-	-	-	-	0.04	±0.2		(15)
$1/K_d^{MT} \times R$	0.03851	±1.711	21.3	±0.353	1.50243	±0.02	2.01	±0.45		(16)
$1/K_d^{MT}$	0.3386	±0.0001	-	-	-	-	1.46	±0.23		(16)
Alternative models:										
Neoantigen load	16.39039	±0.0001	-	-	-	-	1.48	±0.12		(18)
$A \times R$, sum over neoantigens	18.3366	±0.471	38.3	±0.001	10.1531	±29.105	0.21	±0.08		(17)
$A \times R$, alignments at positions 3-8	0.0716	±1.011	22.3	±0.273	0.46591	±0.022	0.94	±0.13		(2)
$A \times R$, negative IEDB assays	0.00091	±0.58	15.8	±0.056	0.45236	±0.001	0.98	±0.03		(2)
$A \times R$, all neoantigens	0.01602	±0.559	34.9	±0.051	0.53933	±0.834	0.13	±0.03		(2)
$A \times R$, average fitness	-	-	26.3	±0.252	1.06061	±0.001	4.03	±0.84	0.04476 *	(2) and (20)

Melanoma, anti-CTLA-4 (ref. 4)	Parameters trained on melanoma, anti-CTLA-4 dataset from ref. 5						Log-rank test			
Models	τ		*a*		*k*		Score		Significance	Equation
	Mean	Std	Mean	Std	Mean	Std	Mean	Std		
$A \times R$	0.02326	±0.479	26	±3.835	3.44101	±0.088	9.1	±0.92	0.00256 ***	(2)
Partial models:										
A	0.1007	±0.0001	-	-	-	-	0.09	±0.68		(13)
R	1.3483	±17.017	21.2	±5.133	5.57735	±27.093	1.17	±0.38		(14)
$K_d^{WT} \times R$	0.00096	±15.302	27	±4.14	5.53499	±0.001	0.53	±0.22		(15)
K_d^{WT}	13.33274	±0.0001	-	-	-	-	1.21	±0.75		(15)
$1/K_d^{MT} \times R$	0.0786	±5.21	22	±0.445	1.4779	±0.065	0.41	±0.65		(16)
$1/K_d^{MT}$	0.06196	±0.0001	-	-	-	-	0.63	±0.61		(16)
Alternative models:										
Neoantigen load	0.10065	±0.0001	-	-	-	-	0.64	±0.21		(18)
$A \times R$, sum over neoantigens	0.08928	±27.215	27	±8.827	5.01647	±34.399	0.09	±0.53		(17)
$A \times R$, alignments at positions 3-8	1.82771	±11.104	24.9	±8.066	8.08455	±1.826	5.33	±1.05	0.021 *	(2)
$A \times R$, negative IEDB assays	0.16414	±10.716	11.7	±0.768	0.89312	±0.164	1.83	±0.97		(2)
$A \times R$, all neoantigens	0.00772	±23.665	25.7	±7.145	7.16555	±0.834	2.63	±0.92		(2)
$A \times R$, average fitness	-	-	26	±3.158	3.34043	±0.001	8.03	±0.92	0.00459 ***	(2) and (20)

```
#
# fitness model parameters
a=26.
k=4.86936
```

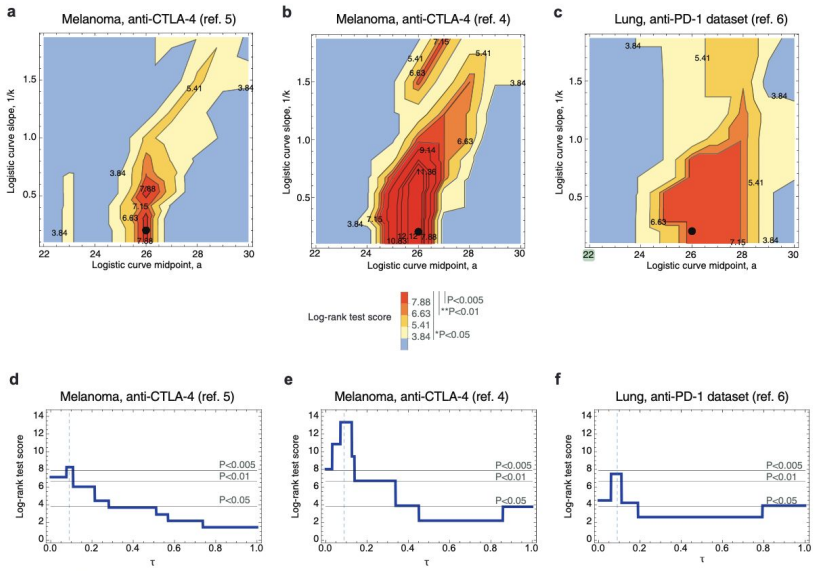
Extremely sensitive to IEDB entries



Extended Data Figure 5 | Effect of IEDB sequence content on predictive power of the neoantigen fitness model. Predictions were performed using subsampled IEDB epitope sequences, with subsampling rate varying between 0.1 and 0.9. For each rate, 10,000 iterations were performed to obtain a distribution of log-rank test scores. The violin plots represent the data density at a given value on the vertical axis ($n = 10{,}000$). Solid black lines mark the log-rank test score of the prediction for the full set of epitope sequences and grey thick lines mark the median scores of the subsampled data. **a–c**, Subsampling of the original set of IEDB sequences, supported by positive T-cell assays, shows that the quality of predictions decreases with subsampling rate. Prediction quality is more robust in the datasets from refs 4, 6. **d–f**, The analogous subsampling procedure was repeated for IEDB sequences that were not supported by positive T-cell assays. For the datasets from refs 4, 5, model performance is substantially decreased.

They tell you it's overfit

We further performed a joint optimization of the cumulative log-rank test score of the three cohorts, obtaining a single set of parameters with predictions that are highly stable around these values (Extended Data Fig. 3). The alignment threshold parameter is consistently set to 26 (Extended Data Table 1), which in our datasets is obtained by alignments of a mean length of 6.8 amino acids, just above the average



Extended Data Figure 3 | Survival analysis score landscape as a function of model parameters. a–c, The landscape of log-rank test scores as the function of the parameters of the TCR-binding model (a and $1/k$), shown for the consistent choice of $\tau = 0.09$; colours represent the significance level of the long-rank test. The regions of high scores are similar across all three datasets. The point corresponding to consistent parameters ($a = 26$ and $k = 4.87$) is marked by a black dot in each plot. **d–f,** log-rank score for the fitness model at consistent binding-function parameters, plotted as a function of τ. Dashed vertical lines are at $\tau = 0.09$, thin solid lines mark the score values corresponding to significance of $P = 0.05$, $P = 0.01$ and $P = 0.005$ ($n = 103$ (**a**, **d**), $n = 64$ (**b**, **e**), $n = 34$ (**c**, **f**)).

❁ *Fin* ❁