Unraveling the Power of NAP-CNB's Machine Learning-enhanced Tumor Neoantigen Prediction

Almudena Méndez-Pérez[1], Andrés M. Acosta-Moreno[1], Carlos Wert-Carvajal[1,2], Pilar Ballesteros-Cuartero[1], Rubén Sánchez-García[1,3], José R Macías[1], Rebeca Sanz-Pamplona[4,5], Ramon Alemany[6], Carlos Óscar S. Sorzano[1], Arrate Muñoz-Barrutia[2], Esteban Veiga[1*].

•1. Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049, Madrid, Spain
•2. Departamento de Bioingenieria, Universidad Carlos III de Madrid, 28911, Leganés, Spain
•3. University of Oxford, Department of Statistics & XChem. Oxford, UK.
•4. Catalan Institute of Oncology (ICO), Oncobell Program, Bellvitge Biomedical Research Institute (IDIBELL), 08908, L'Hospitalet de Llobregat, Spain
•5. Centro De Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
•6. Procure Program, Institut Català d'Oncologia- Oncobell Program, Catalan Institute of Oncology (ICO), Oncobell Program, Bellvitge Biomedical Research Institute (IDIBELL), 08908, L'Hospitalet de Llobregat, Spain

* Corresponding author: eveiga@cnb.csic.es

Abstract.

In this study, we present a proof-of-concept classical vaccination experiment that validates the *in silico* identification of tumor neoantigens (TNAs) using a machine learning-based platform called NAP-CNB. Unlike other TNA predictors, NAP-CNB leverages RNAseq data to consider the relative expression of neoantigens in tumors. Our experiments show the efficacy of NAP-CNB. Predicted TNAs elicited potent antitumor responses *in vivo* following classical vaccination protocols. Notably, optimal antitumor activity was observed when targeting the antigen with higher expression in the tumor, which was not the most immunogenic. Additionally, the vaccination combining different neoantigens resulted in vastly improved responses compared to each one individually, showing the worth of multiantigen-based approaches. These findings validate NAP-CNB as an innovative TNA-identification platform and make a substantial contribution to advancing the next generation of personalized immunotherapies

Introduction.

A new window of hope to treat previously intractable tumors is emerging through immunotherapies (1). However, the response rates of these therapies remain low and relapses are common. Moreover, the severe undesired side effects induce many patients to abandon the treatments (2), highlighting the urgent need for more specific novel therapies.

In this regard, the main challenges for most anti-cancer immunotherapies are the identification of tumor-specific antigens (neoantigens) (3) to avoid undesired side effects and the development of multiantigen-based treatments with combined therapies to prevent tumor relapses (4). A possible strategy to accelerate the search for neoantigens and lower the cost of the therapy is to sequence the cancer cells using Next Generation Sequencing (NGS) techniques and find mutations using bioinformatics tools. Finally, it tries to predict which of the mutations will be more likely to cause an immune response i.e. neoantigen prediction. This area is relatively unexplored with only a few algorithms available (5). The need to validate the algorithmic results has already been recognized as one of the critical steps of this approach (6) and this work specifically addresses it. Algorithmic proposals using deep learning have only started to appear and most of them clearly outperform the standard approaches (7,8). Finding neoantigens in every cancer patient will be fundamental for personalized antitumor immunotherapies (9).
We previously developed an easy-to-use platform (NAP-CNB) that allows to identify tumor neoantigens rapidly (10). NAP-CNB predicts putative neoantigens employing exclusively RNA-Seq reads (10). The tool uses a Long Short-Term Memory (LSTM)-based neural network to rank mutations according to their estimated MHC I complex affinity. NAP-CNB harnesses the suitability of recurrent neural networks to sequential problems to offer high accuracy in affinity binding prediction. Hence, NAP-CNB provides an integrated and resource-efficient pipeline for *in silico* classification of MHC I neoepitopes. In contrast with other tools (11–14), NAP-CNB is entirely automatic and freely available online.

We found NAP-CNB to be comparable or superior to other state-of-the-art methods of murine immunogenicity prediction, like NetH2pan (15) or MHCflurry 2.0 (16), in blind benchmarking. NAP-CNB presents an AUC of 95% for H-2K^{b} typings and a high positive predictive value. The results improved with postprocessing which consists of a majority voting method of an ensemble of sequences presenting single amino acid substitutions. Postprocessing offers a more robust scoring by substituting each amino acid for its most similar one and then classifying the ensemble as the most repeated class.

Results & Discussion

In this work, we analyze *in vivo* the TNA predictive capabilities of NAP-CNB platform using the well-known murine B16 F10 melanoma as a model. For that, we synthesized the peptides corresponding to predicted TNA, used them to vaccinate mice, and analyzed the effectivity against tumor development.

In silico analysis of the B16 F10 melanoma gene expression showed several putative TNA (10) with different scores (predicted probability to be a TNA) and distinct gene expression quantified as fragments per kilobase million (FPKM). We chose for peptide synthesis three top-scored TNA peptides (Fig. 1A); *Pnp (low expression), *Adar (very low expression), *Lrrc28 (low expression). The bottom-scored peptide, *Herc6 (high expression), therefore predicted to not induce any immune reaction against the tumor was chosen as a negative control. In addition, and in order to test whether the post-processing process offers some advantage, a top-scored peptide, *WIZ (high expression), revealed after post-processing was also synthesized (Fig. 1A).

Immunogenicity analysis.
To analyze the immunogenicity of the predicted TNAs, we employed immunocompetent C57BL6/J mice as the recipient model. These mice were vaccinated with individual synthetic peptides *Pnp, *Adar, *Lrrc28, *WIZ, and *Herc6, the latter serving as a putative negative control. In all cases, peptides were emulsified in Freund's adjuvant. A second vaccination boost was injected two weeks after the first inoculation. Four weeks after the first inoculation the immunogenicity of the predicted antigens was tested (Fig. 1B).

We assessed specific cellular responses against the predicted TNAs using an enzyme-linked immunosorbent spot (ELISpot) assay (Fig. 1C, D). Compared to the control group immunized with *Herc6, all vaccinated groups displayed notably elevated levels of interferon-gamma (IFN-γ)-secreting cells targeting TNAs. Notably, *WIZ vaccination yielded a comparatively weaker immune response. This experiment further validates the predicted non-immunogenic nature of the *Herc6 mutation.

Furthermore, we evaluated the immunogenic potential of the predicted TNAs through Intracellular Cytokine Staining (ICS) (Fig. 1E). The $CD8^+$ T-cell population from vaccinated mice was activated by presenting the TNAs on splenocytes, and the production of TNFα, perforin, and CD107a was analyzed via flow cytometry. Vaccination with all TNAs significantly induced the expression of at least one of the specified proteins after specific stimulation. Additionally, we included a well-established positive control, TRP2, a well-known tumor-associated antigen that induces potent cellular responses. As predicted, vaccination with *Herc6 did not elicit immunogenic responses.

These data show that silico-predicted TNAs by NAP-CNB induce robust immune responses following immunization in mice.

***In vivo* anti-tumor implantation assays**.
The antitumor immunogenicity of predicted TNAs was further assessed using immunocompetent C57BL6/J mice, vaccinated with each synthetic peptide or with a peptide mixture containing *Pnp, *Adar, and *Lrrc28, in all cases emulsified in Freund's adjuvant. A second vaccination boost was injected two weeks after the first inoculation and two weeks before tumor implantation (B16 F10 melanoma; see experimental scheme in Fig. 2A). Tumor size was monitored over time, as well as overall survival.

The implanted tumor exhibited robust growth in the control group of non-vaccinated animals (PBS; Fig. 2B). 15 days after implantation, all animals in this group displayed tumors with a volume exceeding 100 mm^3. All animals died 22 days after tumor implantation (Fig. 3A, B). Vaccination with the positive control, TRP2, showed immune responses against B16 F10 detected as slower tumor growth (Fig 2C) and incremented survival rates (Fig. 3A). The median of the tumor growth in control unvaccinated mice (PBS) and vaccinated with TRP2 (positive control) are shown to compare the anti-tumor response of animals vaccinated with the indicated peptides. As expected, the immunization with *Herc6, *in silico* predicted as non-immunogenic, did not result in any positive

outcome either in tumor growth (Fig. 2D) or survival rate (Fig. 3A). Vaccination with *Pnp and *LRRC28 elicited irregular antitumor responses (Fig. 2E, G) with some mice showing excellent responses and other behaving similar to the negative control. These vaccinations showed slightly (not significant) incremented survival rates compared to control (Fig. 3B). Vaccination with *Adar (the less expressed TNA; Fig. 1A) did not reduce the tumor growth (Fig. 2F), nor show any survival improvement (Fig. 3B). However, the vaccination with a mixture of the three predicted TNA (Pnp, Adar, Lrrc28) induced a strong anti-tumor response, comparable and even better than that observed with the positive control, TRP2 (Fig. 2H), resulting in a significantly increased survival (Fig. 3A). In the same regard, the response against tumors in animals vaccinated with *Wiz (predicted TNA after post-processing and highly expressed) alone also elicited a powerful antitumor response (Fig. 2I) with considerably improved survival (Fig. 3A).

Additionally, as a measure of the vaccination effectiveness, it is shown the percentage of mice with tumors remained smaller than 100 mm^3, 15 days after implantation (Fig. 2J). Non-vaccinated animals or animals vaccinated with *Herc6 or *Adar show unsuccessful therapies. Vaccination with *Pnp or *Lrrc28 presented a successful rate of 75 and 50% respectively, and the maximum effectiveness, 100 % of animals, was observed in animals vaccinated with TRP2, *Wiz, or a peptide mixture (*Pnp, *Adar, and *Lrrc28).

Together, these data show that pure *in silico* approaches based on machine learning algorithms are able to identify TNAs that induce strong anti-tumor protection, which is the major bottleneck for most immunotherapies. The rapid identification of tumor neoantigens would allow to target/attack tumors non-treatable today and will vastly improve current immunotherapies, representing a giant step forward in the global anticancer fight. In this context, we demonstrate that the algorithms running in NAP-CNB platform effectively identify TNA that can be used as anti-cancer targets. The proof-of-concept experiments presented herein significantly bolster the prospects of translating TNA identification into practical applications for personalized cancer treatments within society.

Our findings also offer valuable insights for shaping future antitumor interventions. From a pragmatic standpoint, prioritizing TNAs with higher expression levels in the tumor, even if they elicit a comparatively weaker immunogenic response, appears to be a more promising approach. Our data confirm that employing a multiantigen therapy, targeting various tumor epitopes simultaneously, holds significant potential in averting immune escape. This underscores the importance of advancing TNA-based immunotherapy treatments against cancer within the framework of personalized medicine.

Figure legends:

Figure 1. Vaccination-induced immune responses. A. Putative B16 TNA identified by using the NAP-CNB platform ranked by scores of peptide sequences for a complete 12mer sequence. The TNA sequence, the mutation exclusive to tumor cells (in red), and gene name are shown. The gene expression is quantified as fragments per kilobase million (FPKM). The TNA score is also indicated. **B**. Scheme of immunization. 2 doses of peptides emulsified in Freund's adjuvant were s.c. injected separated by 14 days. 14 days after the last dose the efficacy of the vaccine was analyzed by ELISPOT and ICS assays. **C,** ELISPOT analysis of IFNγ-producing T-cell effectors from mice vaccinated with the indicated mutated peptides. The upper images show the response of non-vaccinated (non-vac) animals after stimulation with the indicated peptides, the bottom images show the response of the animals vaccinated (vac) with the indicated peptides after restimulation with the same peptides. It is shown duplicates from representative animals. **D**, as in C but showing the mean and sd of 5 mice per group. Asterisks indicate statistically significant differences analyzed by one-way ANOVA (* represent P< 0.05, ** P< 0.005, *** P< 0.0005). **E,** Intracellular Cytokine Staining (ICS) analysis of $CD8^+$ T cells expressing TNFα, Perforin or CD107a from mice vaccinated with the indicated mutated peptides: *Pnp (black triangle), *Adar (black circle), *Lrrc28 (black square), *WIZ (black rhombus), *Herc6 (grey circle) and TRP2 (green circle). The mean of the ratio of vaccinated divided by unvaccinated is shown, as the 95% confidence intervals. Intervals that do not include 1 and are therefore statistically significant are marked with *.

Figure 2. **Antitumor activity of vaccination with predicted TNA**. **A**. Scheme of immunization. 2 doses of peptides emulsified in Freund's adjuvant were s.c. injected separated by 14 days. 14 days after the last dose the B16 F10 melanoma cells were injected s.c. in the mid-right flank of C57BL/6J host mice and the tumor size over time was analyzed using a dial caliper. **B-I.** Tumor growth on non-vaccinated (B) or vaccinated mice with the indicated peptide (C-I) (or peptide mix (H)), monitored every 1–3 days. Each line corresponds to the tumor size in one animal. The median of the tumors in non-vaccinated (PBS) and vaccinated with TRP2 (positive control) are shown as dashed lines in grey and green color respectively. Mice with tumors ≥ 900 mm^3 were sacrificed. Following the rules of our ethical committee animals presenting ulcers were also sacrificed. **J**. Percentage of animals showing tumors with a volume ≤ 100 mm^3 15 days after implantation.

Figure 3. **A,B** Kaplan-Meier survival curves of mice vaccinated with the indicated TNA peptides and challenged with B16 F10 melanoma. Comparison of Survival Curves has been performed using the Log-rank (Mantel-Cox) test. The significant P values comparing the control (PBS) group with each other group are indicated and * represents P< 0.05.

Bibliography

1. Chen, D. S. & Mellman, I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39,** 1–10 (2013).

2. Kalbasi, A. & Ribas, A. Tumour-intrinsic resistance to immune checkpoint blockade. *Nat. Rev. Immunol.* **20,** 25–39 (2020).
3. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348,** 69–74 (2015).
4. Cable, J. *et al.* Frontiers in cancer immunotherapy-a symposium report. *Ann. N. Y. Acad. Sci.* (2020). doi:10.1111/nyas.14526
5. Boegel, S., Castle, J. C., Kodysh, J., O'Donnell, T. & Rubinsteyn, A. Bioinformatic methods for cancer neoantigen prediction. *Prog Mol Biol Transl Sci* **164,** 25–60 (2019).
6. Vitiello, A. & Zanetti, M. Neoantigen prediction and the need for validation. *Nat. Biotechnol.* **35,** 815–817 (2017).
7. Bulik-Sullivan, B. *et al.* Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4313
8. Wu, J. *et al.* DeepHLApan: A Deep Learning Approach for Neoantigen Prediction Considering Both HLA-Peptide Binding and Immunogenicity. *Front. Immunol.* **10,** 2559 (2019).
9. Tran, E., Robbins, P. F. & Rosenberg, S. A. "Final common pathway" of human cancer immunotherapy: targeting random somatic mutations. *Nat. Immunol.* **18,** 255–262 (2017).
10. Wert-Carvajal, C. *et al.* Predicting MHC I restricted T cell epitopes in mice with NAP-CNB, a novel online tool. *Sci. Rep.* **11,** 10780 (2021).
11. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211,** 2231–2248 (2014).
12. Hundal, J. *et al.* pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8,** 11 (2016).
13. Bjerregaard, A.-M., Pedersen, T. K., Marquard, A. M. & Hadrup, S. R. Prediction of neoepitopes from murine sequencing data. *Cancer Immunol. Immunother.* **68,** 159–161 (2019).
14. Hasegawa, T. *et al.* Neoantimon: a multifunctional R package for identification of tumor-specific neoantigens. *Bioinformatics* **36,** 4813–4816 (2020).
15. DeVette, C. I. *et al.* Neth2pan: A computational tool to guide MHC peptide prediction on murine tumors. *Cancer Immunol Res* **6,** 636–644 (2018).
16. O'Donnell, T. J., Rubinsteyn, A. & Laserson, U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst.* **11,** 42–48.e7 (2020).

Methods.

Peptides *PNP (SLITNKVVMEYENLEKANHM), *ADAR (LVPLSQAWTHPPGVVNPDSC), *LRRC28 (EPMFTFVYPTIFPLRETPMA), *HERC6 (SLVKKWRAAKKRKDREGAKR), *WIZ (TASPPPTARMMFSGLATPSL) and *TRP2 (PQIANCSVYDFFVWLHYYSV) were used for mice immunization. These peptides were synthesized at the proteomics unit from CNB-CSIC. The peptides were synthesized using the stepwise solid-phase peptide synthesis method performed on an automated peptide synthesizer (Multipep, Intavis, Köln, Germany). The amino acid polymerization was carried out using the standard Fmoc (N-(9-fluorenyl) methoxycarbonyl) chemistry and PYBOP/n-methilmorpholine as coupling activation reagents. The Fmoc-derivatized amino acid monomers and the preloaded resins used as support were obtained from Merck. Once synthesized the peptides were cleaved from the resin with a standard scavenger-containing trifluoroacetic acid (TFA)-water cleavage solution and precipitated by addition to cold ether. The crude peptides were purified by reverse-phase chromatography on a semi-preparative HPLC system (Jasco, Tokio, Japan) equipped with a C18 reversed-phase column (Scharlab, Barcelona, Spain). A linear gradient from 5% to 60% solvent B (0.05% trifluoroacetic acid (TFA) in 95% acetonitrile) in solvent A (0.05% TFA in water) was applied for 20 min. The chosen fractions were analyzed in a Maldi-tof 4800 mass spectrometer (Applied Biosystems, Framingham, USA) and those containing the peptide were lyophilized. The peptides were then reconstituted to a concentration of 1 mg/ml in sterile phosphate-buffered saline (PBS) and preserved at -80ºC.

Mouse immunization. Equal volumes of immunogens and Freund's adjuvant (Imject™ Freund's Complete Adjuvant -FCA- for the first dose and Imject™ Freund's Incomplete Adjuvant -FIA- for the second dose, from Thermo Scientific™) were mixed with a double-hub needle until a thick emulsion developed. The final immunogen concentration in the mixtures was 50µg/100µL. C57BL6/J mice were divided into eight groups and were subcutaneously (s.c.) immunized in the left flank with PBS (group 1) or with 100µL of the emulsions containing *TRP2 peptide (group 2), *HERC6 peptide (group 3), *PNP peptide (group 4), ADAR peptide (group 5), *LRRC28 peptide (group 6), *PNP, *ADAR and *LRRC28 peptides mixture (group 7), or *WIZ peptide (group 8). Immunizations were performed at day 0 (with CFA) and 14 (with FIA). At day 28, the efficacy of the vaccine was analyzed by ELISpot and ICS assays, and an anti-tumour experiment was performed.

ELISpot assay. The ELISpot assay was used to detect *PNP, *ADAR, *LRRC28, *HERC6 and *WIZ specific IFNγ secreting cells. 96-well nitrocellulose-bottom plates pre-coated with anti-mouse IFNγ monoclonal antibody were purchased from Mabtech. The plates were blocked with RPMI-10% FBS for at least 30 minutes. After spleen processing, 3×10^5 splenocytes per condition were re-stimulated with 1 µg/ml of the corresponding peptide pool or with RPMI-10% FBS. The plates were incubated with the peptides for 48 h. at 37 °C in 5% CO2 atmosphere, washed five times with PBS, and incubated with 1 µg/ml of the biotinylated anti-mouse IFNγ monoclonal antibody R4-6A2 (Mabtech) diluted in PBS-0.5% FCS for 2 h. at room temperature. The plates were then washed 5

times with PBS and a 1:1000 dilution of ALP-conjugated streptavidin (Mabtech) was added. After 1 h. at room temperature, it was washed 5 times with PBS, and finally developed by adding the ready-to-use substrate solution BCIP/NBT-plus (Mabtech). The reaction was stopped by washing the plate with abundant water. Once it was dry, the spots were counted using the ELISpot Reader System - ELR02- plate reader 651 (AID Autoimmun Diagnostika GmbH) with the aid of AID ELISpot reader system software (Vitro).

Intracellular Cytokine Staining (ICS) assay. The different $CD8^+$ T-cell adaptive immune responses were analyzed by ICS as follows. After spleen processing, 4×10^6 fresh splenocytes (depleted of red blood cells) were seeded on M96 plates and stimulated for 12 h. in complete RPMI 1640 medium supplemented with 10% FBS containing 1µl/ml Brefeldin A (BD Biosciences) to inhibit cytokine secretion, 1µl/ml monensin 1X (eBioscience), anti-CD107a–FITC (BD Biosciences), and the peptides *PNP, *ADAR, *LRRC28, *HERC6, *WIZ, or *TRP2 (1 µg/ml). Cells were then washed, stained for surface markers, fixed (eBioscience™ IC Fixation Buffer), permeabilized (eBioscience™ Permeabilization Buffer), and stained intracellularly with the appropriate fluorochromes. The fluorochrome-conjugated antibodies used for functional analyses were CD3-brilliant violet (BV)-510 (BD Biosciences), CD8 allophycocyanin (APC)-EFluor780 (eBioscience), PERFORIN-APC (Biolegend), and TNF-α–Pacific Blue (Biolegend). Cells were acquired with an LSR-II flow cytometer (BD Biosciences). Analyses of the data were performed with the FlowJo software version 10.4.2 (Tree Star).

Cells. B16-F10 melanoma cell line was obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) and maintained in high glucose Dulbecco′s Modified Eagle′s Medium (DMEM, Gibco-Life Technologies) supplemented with 10% fetal bovine serum (FBS). Cell cultures were maintained at 37 °C in a humidified incubator containing 5% CO2.

Anti-tumor experiment. At day 28 after mouse immunization, B16-F10 cells (4×10^5) were s.c. injected into the mid-right flank of C57BL6/J recipient mice. Tumor growth was measured every 2–3 days with a dial caliper, and the volume was determined by ½(Length×Width^2).

Statistical procedures. One-way ANOVA with Dunnett correction for multiple testing was used for ELISpot analysis to establish the differences within the different groups. We analyzed the ICS results by building a bootstrap distribution based on the ratio of the percentage of specific cell types between the vaccinated and unvaccinated groups. Subsequently, we determined the centered 95% confidence interval. For statistical analysis of the overall survival in the anti-tumor experiment, log-rank (Mantel-Cox) statistical test was performed on day 24. The statistical significances are indicated as follows: *, $p<0.05$; **, $p<0.005$; ***, $p<0.001$.

Ethical statement. Female C57BL/6J mice (6-8 weeks old) used for *in vivo* experiments were purchased from The Jackson Laboratory and stored in the animal facility of Centro Nacional de Biotecnología (CNB-CSIC, Madrid). *In vivo* studies were approved by the Ethical Committee of Animal Experimentation (CEEA) of CNB-CSIC and by the competent authority of *Comunidad de Madrid*

(PROEX 041.4/21). Animal procedures were conformed to international guidelines and to Spanish-European law.

Acknowledgments:

The research is supported by grants: PID2020-116393RB-I00, financed by MCIN/ AEI /10.13039/501100011033/ and BFERO2020.04, financed by FERO foundation. A.M.P. is supported by FPU18/03199JMGG from MCIN. We are grateful to the Proteomics and the Animal Facilities of CNB for peptide synthesis and their support.

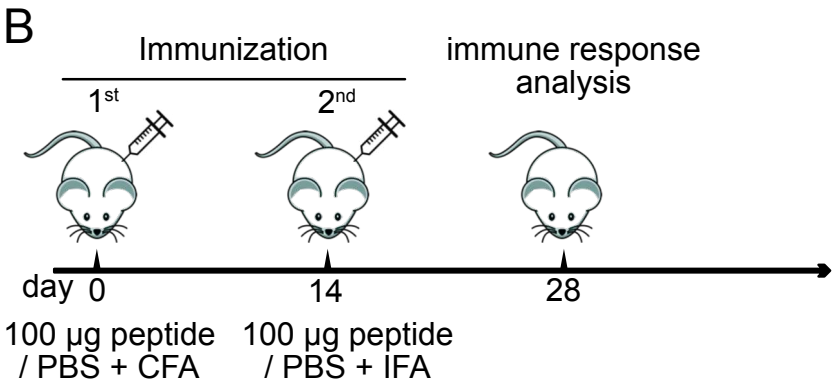
Figure 1

A

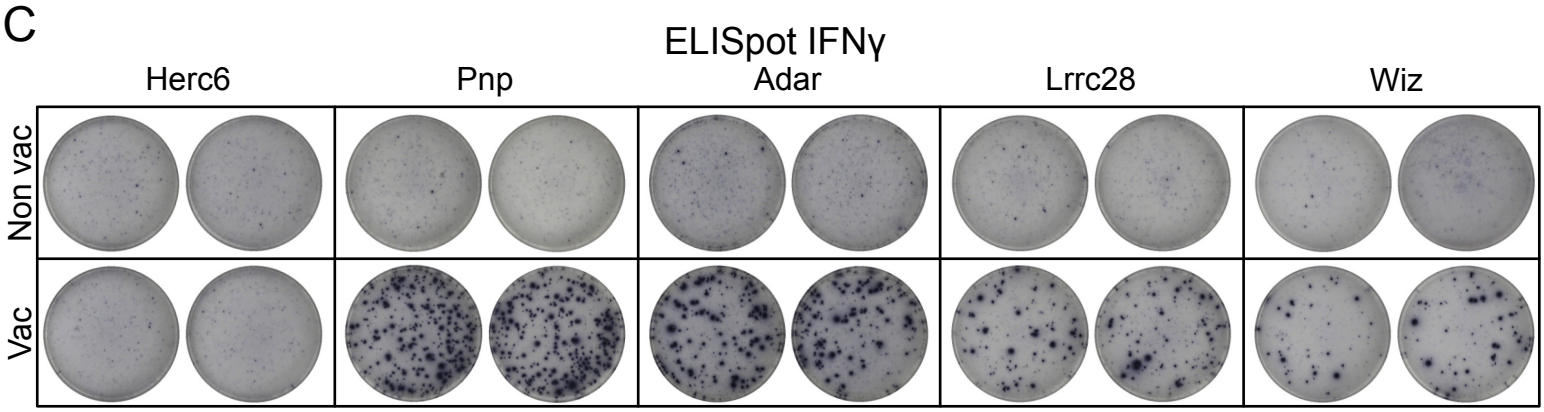
Rank	Sequence	Gene	Score	FPKM
1	NKVVMEYENLEK	MUT Pnp D->E	1.00	3.04
3	SQAWTHPPGVVN	MUT Adar Q->H	1.00	0.00
4	TFVYPTIFPLRE	MUT Lrrc28 K->T	1.00	0.94
…				
34	PPTARMMFSGLA	MUT Wiz K->M	0.03	16.70
393	KWRAAKKRKDRE	MUT Herc6N->K	0.00	10.93

Rank after post processing	Sequence	Gene	Score	FPKM
5	PPTARMMFSGLA	MUT Wiz K->M	1.00	16.70

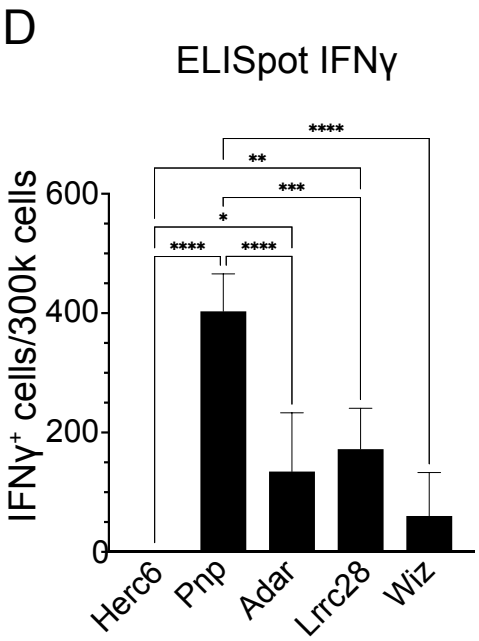
B



C



D



E

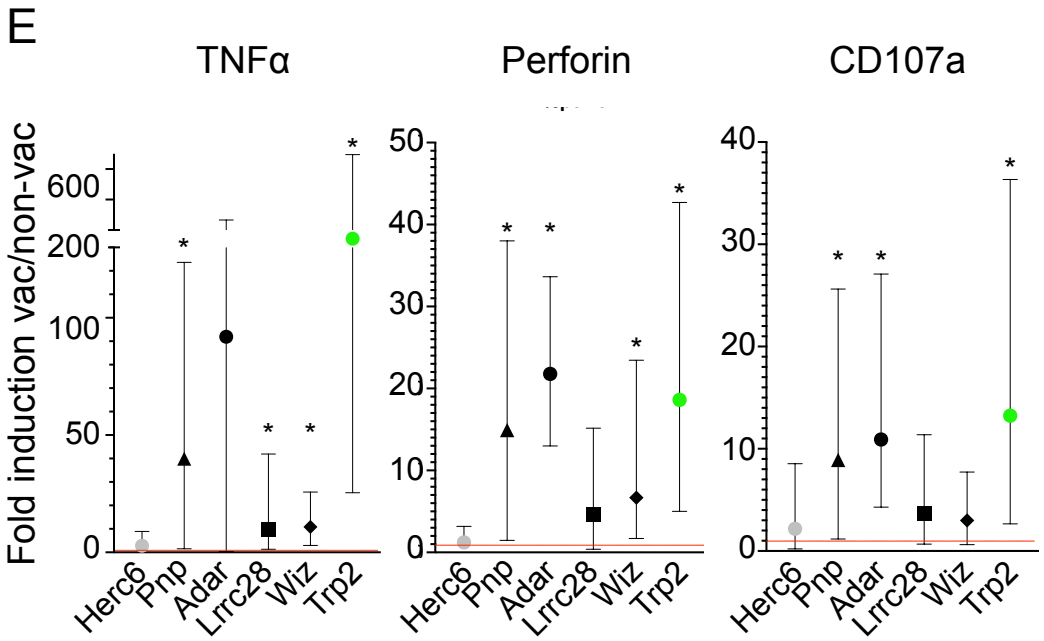


Figure 2

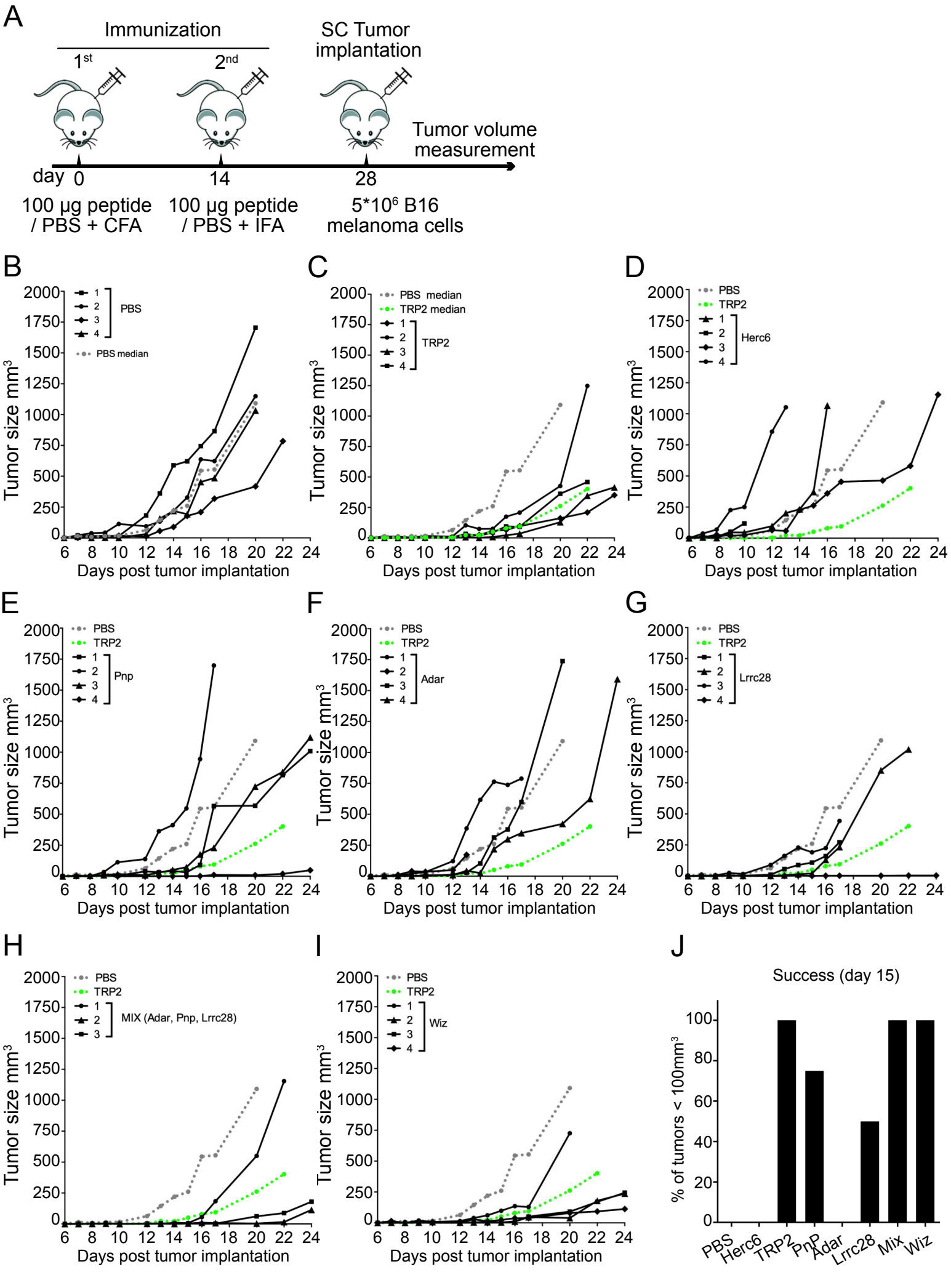


Figure 3