



Universiteit
Leiden

The Netherlands

The repercussions of recognition: imprints of T cells on the tumor microenvironment

Slagter, M.

Citation

Slagter, M. (2025, September 23). *The repercussions of recognition: imprints of T cells on the tumor microenvironment*. Retrieved from <https://hdl.handle.net/1887/4261507>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4261507>

Note: To cite this publication please use the final published version (if applicable).

The Repercussions of Recognition

Imprints of T cells on the Tumor Microenvironment

Maarten Slagter

Cover design by Maarten Slagter.

The repercussions of T cell recognition can reach far beyond the antigen-presenting cell. The titan Atlas, from Greek mythology, represents an HLA molecule, presenting a cytosolic peptide to the extracellular world. The Greek pillars, atypically bent and representing a T cell receptor (TCR), rest heavily on the Atlas's shoulders. The pressure on his feet exerted by the TCR causes the ground to ripple far beyond Atlas's position.

Code to simulate the waves is inspired by

<https://www.ixm-ibrahim.com/explanations/simulating-water-ripples> and available at <https://github.com/slagtermaarten/thesis-waves>.

The depiction of Atlas was generated by Microsoft Copilot and manually shaded. The font used on the cover is Yu Gothic.

The work presented in this thesis has been supported by EU Horizon 2020 project APERIM: Advanced bioinformatics platform for PERsonalised cancer IMmunotherapy (grant agreement ID: 633592).

Printed by: Gildeprint

ISBN: 978-94-6496-451-6

Typesetting of the main text was done using \LaTeX . The font used for the book is Cochineal. Printing of this thesis was financially supported by the NKI-AvL.

Copyright © 2025 *Maarten Slagter*. All rights reserved.

No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior permission of the author and the publisher holding the copyright of the articles.

The Repercussions of Recognition

Imprints of T cells on the Tumor Microenvironment

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op dinsdag 23 september 2025

klokke 16:00 uur

door

Maarten Slagter

geboren te Amsterdam

in 1987

Promotiecomissie

Promotors:	Prof. dr. T. N. M. Schumacher	
	Prof. dr. L. F. A. Wessels	Technische Universiteit Delft
Overige leden:	Prof. dr. R. de Boer	Universiteit Utrecht
	Prof. dr. J. Borghans	Universiteit Utrecht
	Dr. M. Kok	het Nederlands Kanker Instituut
	Dr. A. Mahfouz	
	Prof. dr. K. Visser	

CONTENTS

1	Scope of this thesis	1
I	The clinical utility of T cells in cancer	9
2	Benchmarking the foreignness of human malignancies	11
3	Immune induction strategies in metastatic triple negative breast cancer to enhance the sensitivity to PD-1 blockade: the TONIC-trial	43
II	The effects of T cells on the tumor microenvironment	93
4	Lack of detectable neoantigen depletion in treatment-naïve cancers	95
5	Distinct spatiotemporal dynamics of CD8⁺ T cell derived cytokines in the tumor microenvironment	155
6	Discussion	195

Appendix A	Summary	211
Appendix B	Samenvatting	215
Appendix C	Acknowledgments	219
Appendix D	Curriculum vitae	221
Appendix E	List of Publications	223

SCOPE OF THIS THESIS

This thesis is about immune cells of the adaptive immune system called T cells, how they interact with the microenvironment of tumors, which types of tumors are sufficiently visible to the T cell based immune system and how we can better boost T cell activity to treat patients with cancer. Put differently, the results in this thesis explore the effects of antigen-specific T cells on the tumor masses that they have homed to. Summarizing part of this work: a T cell's impact reaches beyond the directly adjacent antigen-expressing cell, but not detectably so in all manners we expected it to.

But first, what exactly are these immune cells called T cells? In the evolution of life, multicellular organisms have had to develop immune systems in order to fight off pathogens (such as bacteria and viruses) and to maintain homeostasis. The immune system of jawed vertebrates, the phylum that contains humans and mice as well 99% of all other modern vertebrates, consists of a plethora of different cell types, each with their own role to play in supporting immunity. The classical way to organize the ontology of immune cells is to group them into the *innate* and *adaptive* branches, but we now know of many cell types that exhibit characteristics of both branches and so the reality is not quite as binary. Sketching characteristics of the innate and adaptive branches in broad strokes one could say that the innate immune system offers a first line of defense, with rapid responses to a broad group of pathogens based on general molecular patterns. In contrast, the adaptive immune system

has the distinct potential for higher specificity (tailoring responses to individual pathogens), and without a requirement for recurring molecular patterns, albeit at a slower pace than the innate immune system. Importantly, the adaptive immune system has the capability to form immunological memory, such that it can respond with higher agility when a pathogen is encountered again. The adaptive system is also where we find the protagonist of this thesis: the T lymphocyte or T cell.

A distinguishing feature of T cells is that they can survey the entire proteome (including inner proteins) of an organism's own (i.e., host) cells, to try and discriminate between healthy cells and those that may have been compromised by a pathogen. To do this, T cells carry a protein complex called the T cell receptor (TCR) on their cell surface, which provides its molecular specificity. The TCR is capable of interacting with short peptides, derived from cellular proteins, that are presented on the cell surface by major histocompatibility complex (MHC) molecules. Together, they form the so called peptide MHC complexes (pMHCs). Essentially all host cells present pMHCs on their cell surface as a way of offering 'identification' to the T cell based immune system and can be instructed to further increase this activity by different kinds of molecular 'messengers' that put the immune system on alert. As highly specific sentinels, T cells migrate through the body and use their TCR to scan the pMHCs they encounter. Upon encountering a fitting 'match', the pMHC-bound TCR activates an intracellular signal transduction pathway within the T cell, consisting of a series of biochemical events mediated by associated co-receptors, adaptor molecules, and activated transcription factors, that will trigger various processes to neutralize the cell presenting the matching pMHC. Via this mechanism, T cells can track down virus (e.g. influenza, corona) infected cells. More specifically, infection of a host cell by a virus typically leads to expression of viral proteins, which can result in MHC-presentation of 'viral' peptides and T cell recognition of these infected cells.

However, the killing capacity of T cells reaches beyond virally infected cells. It has become abundantly clear that T cells can also detect and kill cancer cells, with large therapeutic potential for patients with cancer. Early evidence in favor of a potential role of T cells in the control of cancer came from the observation that, for a large variety of cancers, the intratumoral infiltration of CD8⁺ T cells correlates with positive prognosis^{1,2}. Direct evidence for the anti-tumor potential of T cells came from data demonstrating that the administration of antibodies that target T cell inhibitory receptors, such as CTLA-4 and PD-1 (also known as T cell checkpoint blockade), shows clinical benefit in many different forms of cancer^{3,4}. By the same token, the infusion of ex vivo-expanded autologous tumor-infiltrating lym-

phocytes (TILs) can induce clinically meaningful responses in melanoma patients⁵, even when refractory to anti-PD-1 treatment⁶. Finally, early data on personalized (mRNA) vaccination, aiming to boost anti-tumor T cell reactivity and expand naturally occurring T cell clonotypes, points towards a potential clinical benefit for patients with melanoma⁷⁻⁹ and pancreatic ductal adenocarcinoma¹⁰.

As most human cancers are not associated with pathogen infection, how does the T cell-based immune system see cancer cells as foreign? Malignant transformation of cells depends on accumulation of DNA damage, which is a double-edged sword to an individual cancer cell. On the one hand, DNA damage may confer a fitness advantage over neighboring wild type (unmutated) cells, by enabling the acquisition of cancer hallmarks (e.g., increased proliferative capacity, resistance to cytostatic environmental cues). On the other hand, mutationally altered proteins can lead to the presentation of non-self *neoantigens*, which allow the T cell based immune system to identify the cancer cell as 'foreign'^{11,12}. T cells can respond to the neoantigens that arise as a consequence of such genomic alterations^{13,14} and this is likely to, at least in part explain, the clinical activity of both T cell checkpoint blockade⁴ and adoptive TIL therapy¹⁴.

Part I - The clinical utility of T cell therapies in cancer

The first part of this thesis is of a translational nature, with a relatively direct applicability to clinical care for patients with cancer.

My long PhD period spanned a pivotal time in the field of T cell targeting immunotherapies for cancer. Back in 2015 (the start of my PhD), highly encouraging data for the clinical utility of T cell checkpoint blockade started to emerge in melanoma and non-small cell lung cancer¹⁵, tumor types that tend to carry a high mutational burden, and thereby neoantigen burden. If mutation derived neoantigens are indeed critical for T cell recognition of cancer, then one can wonder which other cancer types are sufficiently rich in this antigen class. There was a strong desire to get an early idea of the clinical utility of immunotherapies such as T cell checkpoint blockade beyond the aforementioned highly mutated tumor types. Leveraging large cancer sequencing projects like the Cancer Genome Atlas (TCGA) and the International Consortium for Cancer Genomes (ICGC), that have provided a molecular characterization of thousands of cancer samples, we addressed this need in **Chapter 2**. To learn what numbers of antigens may be sufficient for clinically relevant T cell activity, we first assessed the antigen load of 'benchmark' viruses, of which the controllability by

the T cell-based immune system is well documented. We then applied a custom neoantigen prediction pipeline, routinely used for the prioritization of neoantigens for experimental screening, to assess the different forms of DNA damage for their neoantigen generating potential and the cumulative predicted neoantigen load for each of the samples. Encouragingly, we found ~50% of the assessed cancer samples to be richer in predicted (neo)antigens than one of the benchmark viruses, suggesting potentially widespread applicability of T cell engaging therapies for cancer.

In **Chapter 2**, one subgroup of breast cancer patients, those of the basal subtype, appeared to have especially high predicted neoantigen loads. This basal subtype in the PAM50 subtyping system is largely similar to the triple negative breast cancer (TNBC) subtype of the molecular breast cancer subtyping system. TNBC typically affects younger women and is highly aggressive¹⁶. Until recently, the median survival for patients with metastatic disease was a meager 12-18 months^{17,18}. Unlike other forms of breast cancer, TNBC is not driven by the hormones progesterone and estrogen, nor by amplification of the oncogene *HER2*. Instead, TNBC is typically defective in the apoptosis regulator P53, and less frequently in the DNA damage repair proteins BRCA1/2, contributing to TNBC's high mutational burden relative to other breast cancer subtypes. Based on data obtained in e.g., melanoma and non-small cell lung cancer, the high mutational burden and concomitant neoantigen load could offer a window of opportunity for T cell recognition. In **Chapter 3**, we describe the TONIC-study that evaluated the clinical utility of T cell checkpoint blockade, a treatment form that hinges on the availability of tumor specific antigens that can be recognized by T cells. In this phase II signal finding study, all 67 patients received nivolumab, a therapeutic antibody that antagonizes a regulatory protein on the cell surface of T cells (programmed death receptor 1, PD1). To try and elevate the intratumoral levels and/or activity of T cells prior to nivolumab administration, different induction treatments were evaluated: radiation therapy (3 times 8 Gray), oral cyclophosphamide, cisplatin, low-dose doxorubicin or a waiting period of two weeks. **Chapter 3** details an extensive molecular characterization of longitudinally acquired biopsies, aimed at comparing the induction treatments and gaining an understanding of their mechanism of action, as well as on identifying molecular markers of clinical benefit.

Part II - The effects of T cells on the tumor microenvironment

The second part of my thesis is of a more fundamental nature, primarily aimed at expanding our understanding of T cell biology in the context of cancer.

In **Chapter 5**, we investigate the way T cells communicate with other cells in the tumor microenvironment. Specifically, we report on the spreading behaviour of the T cell emitted cytokines IFN- γ and TNF- α . For this, we leverage the cytokine responsive behaviour of numerous endogenous genes and single cell RNA-sequencing of antigen-negative cells, which cannot directly be engaged by T cells. We found IFN- γ to spread beyond the antigen-presenting cell, consistent with earlier microscopy-based work^{19,20}, whereas TNF- α 's activity is - surprisingly - more confined. We also acquired evidence for the notion that TGF- β -sensing is lowered in IFN- γ -experienced cells, potentially due to IFN- γ -instructed lowering of TGF- β secretion in neighboring immune cells.

In **Chapter 4**, we revisit the neoantigen predictions of **Chapter 2** to study the ramifications of the selective pressure exerted by the T cell based immune system on developing cancers. It has long been hypothesized that, through their killing activity, T cells may exert evolutionary selective pressure on cancers²¹, even before clinical manifestation and intervention. Such 'immune surveillance' is supported by murine models of cancer. Immunodeficiencies in mice increase tumor incidence and susceptibility to transplanted or chemical carcinogen-induced tumors²². For humans, (direct) evidence of immunoediting is more elusive. Some hints of its existence can be gleaned from the fact that immunosuppression, either due to AIDS²³ or purposefully induced to facilitate organ transplantation²⁴, is associated with a higher incidence of especially virus-induced cancers. However, these associations can also be attributed to decreased immunity against these (oncogenic) viruses, rather than a decreased ability to clean up nascent tumors.

Nevertheless, an important component of cancer development may be the evasion of immunity²⁵ and a multitude of mechanisms to this effect have been identified. For instance, tumors may lose components of the IFN- γ -signalling pathway²⁶, thereby conferring 'selective deafness' to IFN- γ , a central cytokine emitted by activated T cells. Tumors can also undergo genetic loss of the MHC locus²⁷, prohibiting T cell recognition via the T cell receptor. A third way to evade T cell immunity is to genetically lose DNA mutations associated with neoantigens²⁸ or transcriptionally silence expression of neoantigen coding genes^{29,30}, but the (bioinformatic) detectability of this type of immune editing in treatment-naïve cancer is still controversial³¹. In **Chapter 4** we leveraged the great statistical power that the

large availability of patient samples should offer, to try and robustly detect neoantigen depletion. We could not find a convincing signal of depletion, despite having corrected for a large series of potentially confounding processes. This result is likely to be in large part due to the high false positive rate that currently still plagues neoantigen prediction³², a key tool in these analyses. As such, the statistical power of these analyses is limited and they should not be interpreted as fully ruling out some degree of sculpting of the neoantigen repertoire.

Finally, in **Chapter 6**, I first discuss how to build upon the methodology we developed for transcriptome-based cytokine exposure inference in **Chapter 5**. Due to the many cytokine-responsive genes and unique patterns with which different cytokines may up- or downregulate gene expression, the potential for this approach reaches far beyond 'just' the handful of cytokines we studied in **Chapter 5**, while its applicability also continues to rise with the developing revolution of *spatial* single cell sequencing. To enable the inference of exposure to stimuli (e.g., cytokines) from RNA-seq in a highly multiplexed fashion, algorithmic innovation will be of critical importance. In this chapter, I describe potential issues in applying transcriptome-based cytokine inference and how these can be surmounted with improved computational frameworks.

References

1. Galon, J. *et al.* Type, Density, and Location of Immune Cells within Human Colorectal Tumors Predict Clinical Outcome. *Science (New York, N.Y.)* **313**, 1960–1964. ISSN: 1095-9203 (Sept. 2006).
2. Fridman, W. H., Pagès, F., Sautès-Fridman, C. & Galon, J. The Immune Contexture in Human Tumours: Impact on Clinical Outcome. *Nature Reviews Cancer* **12**, 298–306. ISSN: 1474-175X, 1474-1768. (2012) (Apr. 2012).
3. Ribas, A. & Wolchok, J. D. Cancer Immunotherapy Using Checkpoint Blockade. *Science (New York, N.Y.)* **359**, 1350–1355. ISSN: 1095-9203 (Mar. 2018).
4. Morad, G., Helmink, B. A., Sharma, P. & Wargo, J. A. Hallmarks of Response, Resistance, and Toxicity to Immune Checkpoint Blockade. *Cell* **184**, 5309–5337. ISSN: 1097-4172 (Oct. 2021).
5. Hinrichs, C. S. & Rosenberg, S. A. Exploiting the Curative Potential of Adoptive T-cell Therapy for Cancer. *Immunological Reviews* **257**, 56–71. ISSN: 1600-065X. (2014) (2014).
6. Rohaan M. W. *et al.* Tumor-Infiltrating Lymphocyte Therapy or Ipilimumab in Advanced Melanoma. *New England Journal of Medicine* **387**, 2113–2125. (2024) (Dec. 2022).

7. Ott, P. A. *et al.* An Immunogenic Personal Neoantigen Vaccine for Patients with Melanoma. *Nature* **547**, 217–221. ISSN: 0028-0836. (2017) (July 2017).
8. Sahin, U. *et al.* Personalized RNA Mutanome Vaccines Mobilize Poly-Specific Therapeutic Immunity against Cancer. *Nature* **547**, 222–226. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). PMID: 28678784 (2017).
9. Weber, J. S. *et al.* Individualised Neoantigen Therapy mRNA-4157 (V940) plus Pembrolizumab versus Pembrolizumab Monotherapy in Resected Melanoma (KEYNOTE-942): A Randomised, Phase 2b Study. *Lancet (London, England)* **403**, 632–644. ISSN: 1474-547X (Feb. 2024).
10. Rojas, L. A. *et al.* Personalized RNA Neoantigen Vaccines Stimulate T Cells in Pancreatic Cancer. *Nature* **618**, 144–150. ISSN: 1476-4687. (2023) (June 2023).
11. Schumacher, T. N. & Schreiber, R. D. Neoantigens in Cancer Immunotherapy. *Science (New York, N.Y.)* **348**, 69–74. ISSN: 1095-9203 (Apr. 2015).
12. Schumacher, T. N., Scheper, W. & Kvistborg, P. Cancer Neoantigens. *Annual Review of Immunology* **37**, 173–200. ISSN: 0732-0582, 1545-3278. (2024) (Apr. 2019).
13. van Rooij, N. *et al.* Tumor Exome Analysis Reveals Neoantigen-Specific T-cell Reactivity in an Ipilimumab-Responsive Melanoma. *J Clin Oncol* **31**, e439–42. ISSN: 1527-7755 (Electronic) 0732-183X (Linking). PMID: 24043743 (2013).
14. Robbins, P. F. *et al.* Mining Exomic Sequencing Data to Identify Mutated Antigens Recognized by Adoptively Transferred Tumor-Reactive T Cells. *Nature Medicine* **19**, 747–752. ISSN: 1546-170X (June 2013).
15. Postow, M. A., Callahan, M. K. & Wolchok, J. D. Immune Checkpoint Blockade in Cancer Therapy. *Journal of Clinical Oncology* **33**, 1974–1982. ISSN: 0732-183X. (2024) (June 2015).
16. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-Negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease. *Nature Reviews. Clinical Oncology* **13**, 674–690. ISSN: 1759-4782 (Nov. 2016).
17. Kassam, F. *et al.* Survival Outcomes for Patients with Metastatic Triple-Negative Breast Cancer: Implications for Clinical Practice and Trial Design. *Clinical Breast Cancer* **9**, 29–33. ISSN: 1526-8209 (Feb. 2009).
18. Skinner, K. E., Haiderali, A., Huang, M. & Schwartzberg, L. S. Real-World Effectiveness Outcomes in Patients Diagnosed with Metastatic Triple-Negative Breast Cancer. *Future Oncology (London, England)* **17**, 931–941. ISSN: 1744-8301 (Mar. 2021).
19. Thibaut, R. *et al.* Bystander IFN- γ Activity Promotes Widespread and Sustained Cytokine Signaling Altering the Tumor Microenvironment. *Nature Cancer* **1**, 302–314. ISSN: 2662-1347. (2024) (Mar. 2020).
20. Hoekstra, M. E. *et al.* Long-Distance Modulation of Bystander Tumor Cells by CD8+ T-cell-secreted IFN- γ . *Nature Cancer* **1**, 291–301. ISSN: 2662-1347 (Mar. 2020).

-
21. Ribatti, D. The Concept of Immune Surveillance against Tumors: The First Theories. *Oncotarget* **8**, 7175–7180. ISSN: 1949-2553. (2024) (Oct. 2016).
 22. Schreiber, T. H. & Podack, E. R. A Critical Analysis of the Tumour Immunosurveillance Controversy for 3-MCA-induced Sarcomas. *British Journal of Cancer* **101**, 381–386. ISSN: 0007-0920, 1532-1827. (2024) (Aug. 2009).
 23. Yarchoan, R. & Uldrick, T. S. HIV-Associated Cancers and Related Diseases. *The New England journal of medicine* **378**, 1029–1041. ISSN: 0028-4793. (2024) (Mar. 2018).
 24. Engels, E. A. *et al.* Spectrum of Cancer Risk among US Solid Organ Transplant Recipients. *JAMA* **306**, 1891–1901. ISSN: 1538-3598 (Nov. 2011).
 25. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674. ISSN: 0092-8674. (2024) (Mar. 2011).
 26. Zaretsky Jesse M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *New England Journal of Medicine* **375**, 819–829. (2024) (2016).
 27. Martínez-Jiménez, F. *et al.* Genetic Immune Escape Landscape in Primary and Metastatic Cancer. *Nature Genetics* **55**, 820–831. ISSN: 1546-1718. (2024) (May 2023).
 28. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacoheh, N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* **160**, 48–61. ISSN: 10974172 (2015).
 29. Verdegaal, E. M. E. *et al.* Neoantigen Landscape Dynamics during Human Melanoma–T Cell Interactions. *Nature* **536**, 91–95. ISSN: 1476-4687. (2024) (Aug. 2016).
 30. Rosenthal, R. *et al.* Neoantigen-Directed Immune Escape in Lung Cancer Evolution. *Nature* **567**, 479–485. ISSN: 1476-4687. (2024) (Mar. 2019).
 31. Van den Eynden, J., Jiménez-Sánchez, A., Miller, M. L. & Larsson, E. Lack of Detectable Neoantigen Depletion Signals in the Untreated Cancer Genome. *Nature Genetics* **51**, 1741–1748. ISSN: 1061-4036, 1546-1718. (2023) (Dec. 2019).
 32. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818–834.e13. ISSN: 00928674. (2023) (Oct. 2020).

Part I

The clinical utility of T cells in cancer

BENCHMARKING THE FOREIGNNESS OF HUMAN MALIGNANCIES

Maarten Slagter^{*}, Lorenzo F. Fanchi^{*}, Marit M. van Buuren, Jorg J.A. Calis, Philip C. Schouten, Andrew Menzies, Roel G.W. Verhaak, Arno Velds, Ron M. Kerkhoven, Gergana Bounova, Sabine C. Linn, Hendrik Veelken, Michael R. Stratton, Ludmil B. Alexandrov, Lodewyk F.A. Wessels and Ton N. Schumacher^{*} These authors contributed equally

Cell Press preprint server: 10.2139/ssrn.3279415

Abstract

Mutational load varies widely between and within malignancies and has been used as a proxy for the immunological foreignness of human cancers. However, without well-defined reference points it is difficult to determine which human tumors can be considered sufficiently foreign to the T-cell-based immune system. We established a neoantigen prediction pipeline that processes single nucleotide variants, indels and structural variants and established its precision in identifying T-cell-recognized antigens. We subsequently used this pipeline to benchmark the immunological foreignness of human cancers against that of human pathogens for which T-cell control has been established. We demonstrate that up to 50% of tumors, spanning 25 sites of origin, are more foreign than these pathogen benchmarks, due to the presentation of foreign antigens they have accumulated. These data suggest that enhancing the activity of the endogenous tumor-specific T-cell compartment through immunotherapeutic strategies may be of value for a large fraction of human cancers.

Introduction

The ability of the T-cell-based immune system to specifically recognize and destroy human cancers has attracted strong attention. Antibodies against the T-cell checkpoint molecules CTLA-4, PD-1, and its ligand PD-L1, can induce tumor regression in a range of human malignancies^{1–7}. In addition, both infusion of autologous *ex vivo*-expanded tumor-infiltrating lymphocytes (TIL) and TCR gene-modified peripheral blood lymphocytes have shown activity in melanoma⁸ and HPV-associated cancers^{9,10}. The anti-tumor activity of these therapies is at least partially mediated by CD8⁺ T cells, as suggested by the predictive value of pre-treatment intratumoral CD8⁺ infiltrates¹¹ and as shown by the clinical activity of purified CD8⁺ TIL¹². Several lines of evidence indicate that T-cell recognition of neoantigens that are formed as a consequence of somatic DNA damage are a major driving force behind the activities of these therapies. First, T-cell responses against neoantigens are observed in a large fraction of patients with highly mutated tumors^{6,13} and can be boosted by immunotherapy^{6,14}. Second, activity of both CTLA-4 and PD-1/PD-L1 blockade is preferentially observed in tumors with above average mutational burden in a number of tumor types^{15–18}. Third, and most directly, case reports that describe the infusion of T-cell products with high levels of neoantigen reactivity showed evidence of clinical activity and associated selective pressure against neoantigenic mutations^{19–21}.

The availability of genomic information on different human cancers has inspired efforts to estimate the immunological foreignness of these cancers through *in silico* predictions. In a landmark paper, Allison and Vogelstein postulated that many human tumors should carry neoantigens²². Subsequent work has similarly used genomic data to predict the presumed neoantigen burden of different human cancers^{23,24}, in some cases in combination with clinical response to checkpoint inhibitors^{15,25,26}. While these efforts have been very valuable to describe the relative number of predicted neoantigens in different human cancers, the data have not been well-suited to draw quantitative conclusions as a grounded reference point informing on the number of epitopes sufficient to allow the formation of protective T-cell responses has been lacking.

To advance our understanding of immunological foreignness of different human cancers, we first developed an epitope prediction pipeline (Neolution) and assessed its performance using a set of experimentally identified T-cell recognized HIV epitopes. We subsequently used this pipeline to predict neoantigens for a total of 7,290 tumors obtained from the Cancer Genome Atlas (TCGA), the International Cancer Genomics Consortium (ICGC) and the Multiple Myeloma Foundation (CoMMpass) repositories, taking into account the consequences of single nucleotide variants (SNVs), insertions-deletions (indels) and structural variants, and – where applicable – included other potential sources of immunologically foreign protein sequence encoded by non-germline DNA sequences, such as proteins encoded by oncogenic viruses and the B-cell receptor idotype. In this manner, we transformed the entire pool of genomically novel sequences of individual tumors to the foreign antigen space (FAS), allowing for their direct comparison to a grounded, viral context. To generate such

context, we used the same prediction pipeline to analyze proteins from three pathogens for which T-cell control has been shown to be clinically relevant: the *E6* and *E7* oncogenes from the human papilloma virus (HPV)²⁷, the *LMP-1* and *LMP-2* oncogenes from the Epstein-Barr virus (EBV)²⁸, and the HIV-1 genome^{29,30}.

We demonstrate that a large fraction of human cancers has evolved to express a sufficient number of foreign antigens to allow T-cell recognition. Tumors positive for a viral integration are shown to be especially immunologically foreign. These data provide a strong incentive for the further development of therapeutic strategies that aim to expand (neo-)antigen-specific T-cell reactivity in not only highly, but also modestly mutated human tumor types, especially those of viral aetiology.

Results

Robust estimates of DNA damage- and virus-derived antigen loads

To generate a pan-cancer overview of the foreign antigen space of human tumors, we selected pre-treatment samples from 7,268 patients across 25 tissues of origin, covering 42 tumor subtypes, based on the availability of patient-matched DNA and RNA sequencing data (Figure 2.1, Table for list of patient identifiers and annotations). We reconstructed tumor transcripts using SNVs, indels and, for a subset of patients from whom these data were available, large-scale gene fusion events in order to obtain a set of candidate tumor-specific neoantigens. For both SNVs and indels, candidate peptides whose genomic sequences were affected by multiple mutations were modified to reflect the consequences of all variants (154,731 of 8,818,152 candidate peptides, 1.75%). Conversely, structural variants were analyzed in isolation as these were found not to substantially overlap with focal DNA damage (i.e. missense mutations and indels; only 9 out of 20,140 fusion events affected, Figure S2.1A). NMD-targetability of transcripts was predicted and targeted transcripts and their associated candidate peptides were excluded, where indicated. Finally, candidate neoantigens were annotated with the output of our four-filter epitope prediction pipeline that models the major requirements for (neo-)antigen presentation individually: expression of the mutant DNA sequence, predicted proteasomal processing, predicted HLA-binding, and self-similarity (Figure S2.1A). RNA expression was assessed using sample-matched gene expression levels from the matched tumor, requiring at least one mapped read to be considered expressed, unless indicated otherwise. HLA-binding was assessed with the frequently observed HLA-A*02:01 allele for which binding affinity predictions are most accurate³¹. Using a single HLA-allele rather than patient-matched HLA-alleles allows for uniform prediction accuracy across patients. With any of four other HLA-alleles that are also well-predicted, we did not get substantially different neoantigen loads (Figure S2.1B). From 959,792 non-synonymous DNA mutation events, we predicted 80,234 candidate peptides that passed all four filters.

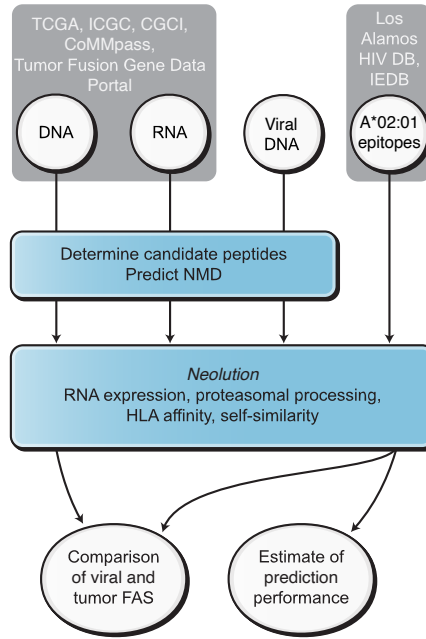


Figure 2.1: Study setup and methodology.

Tumor genome sequencing data and viral genome sequences were used as input for the indicated neo-antigen prediction pipeline, yielding tumoral and viral foreign antigen loads for direct comparison.

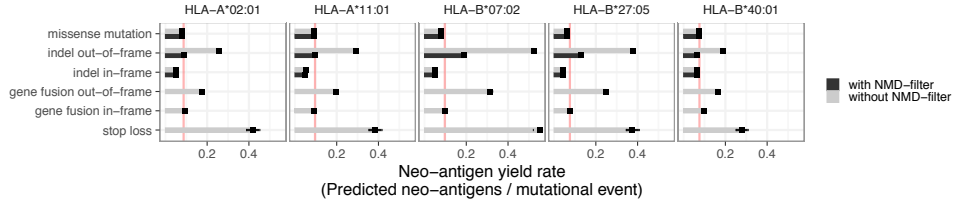
Focal DNA damage dominates the foreign antigen space of most tumors

Previous neoantigen identification efforts have mostly focused on SNVs rather than other classes of DNA damage like indels and gene fusions^{14,23,24}. Frameshifting indels and loss-of-stop mutations can result in open reading frame extensions, where novel coding sequences ultimately become accessible to RNA translation, forming an additional source of neo-epitopes^{23,32}. Similarly, while in-frame gene fusion events yield immunologically novel DNA sequences that are restricted to the fusion breakpoint, out-of-frame fusions additionally lead to the formation of downstream novel open reading frames (ORFs) until an in-frame stop codon is encountered. As can be expected, peptides derived from these novel ORFs were more likely to be dissimilar from the self-ligandome and thus have a higher probability of being T-cell recognized: of all affinity-filter-passing candidate peptides 74.0% (95% CI: [73.5%, 74.5%]) and 61.0% (95% CI: [60.8%, 61.2%]) derived from frameshifting indels and missense mutations, respectively, were classified as dissimilar from self. Next, we determined the propensity of each mutation class to generate neo-epitopes by assessing the ratio between the number of neoantigens contributed by the class and the prevalence of that class, resulting in a mutation-class-specific neoantigen yield rate. As expected, we found DNA damage types to vary significantly

in their neoantigen yield rates (Figure 2.2A) in a manner highly consistent across tumor types (Figure S2.2A) and HLA alleles. Loss-of-stop mutations generated 4.99-fold more neoantigens per mutation than missense mutations. The neoantigen yield rate of frameshifting indels only minimally exceeded that of missense mutations when accounting for NMD (1.1-fold), but yielded 3.1-fold more neoantigens when omitting the NMD-filter (Figure 2.2A). In spite of this observed variation in neoantigen yield rate between DNA damage classes, the absolute contributions of different mutation classes to the predicted antigenome is predominantly determined by their prevalence. Specifically, due to the high abundance of missense mutations, this class dominates the predicted antigenome for all investigated tumor types, despite its relatively modest neoantigen yield rate. This holds true both when NMD is not taken into account (Figure 2.2B, top) and when it is by removing predicted neoantigens derived from transcripts identified as NMD-targeted (Figure 2.2B, bottom).

Many of the mutational processes underlying tumorigenesis (e.g. UV radiation, tobacco smoke and loss of DNA repair mechanisms) are associated with unique patterns of SNVs, termed mutational signatures³³. To assess possible differences in neoantigenicity between mutational processes, we inferred the likelihood for SNVs to have been caused by any of the identified processes and associated signatures for 6,504 tumor samples across 36 tumor types (methods, Table). This calculation was performed for each SNV individually taking into account the signature's tendency to cause the SNV and the total abundance of the signature. We observed small and statistically insignificant differences in HLA-A*02:01 neoantigen yield rates between signatures within tumor types (lowest FDR-adjusted *p*-value is 0.214 for lung squamous cell carcinoma, Figure S2.2B) indicating that mutational processes do not differ in their ability to yield SNV-associated HLA-A*02:01 (or other alleles, data not shown) neoantigens.

A



B

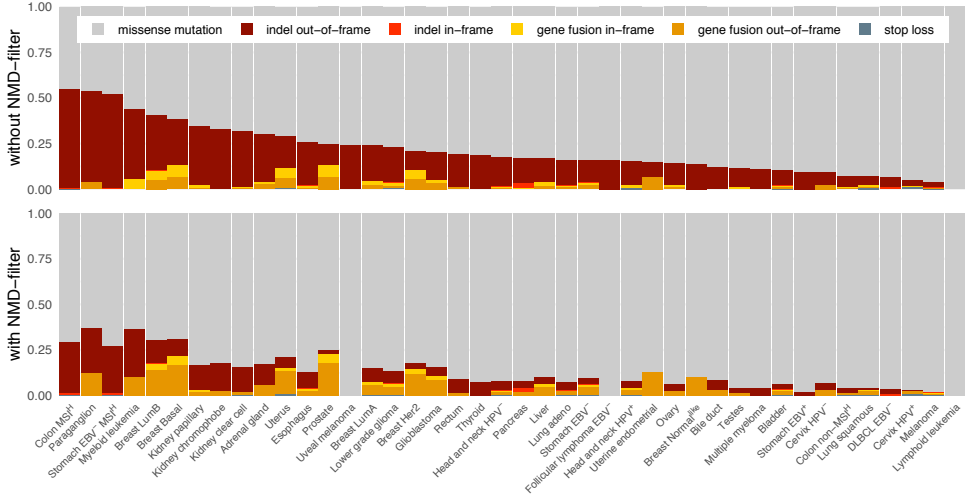


Figure 2.2: Neo-antigen contribution and neo-antigen yield rates of different types of DNA damage

A Neo-antigen yield rates (the mean number of predicted neoantigens per mutation) for five different HLA alleles and different classes of DNA damage. Weighted mean yield rates indicated by squares, surrounding lines denote 95% CIs.

B Proportional contribution of different types of DNA damage to the DNA damage-associated foreign antigen loads across human tumors, both without (top) and with (bottom) the NMD-filter applied. In the bottom panel, gene fusion and stop loss variants were assumed not to be affected by NMD.

Viral benchmarks to compare DNA-damage derived antigen loads with

Prior analyses have demonstrated that the number of predicted neoantigens correlates well with the number of genomic alterations^{6,24,26}. However, in absence of a benchmark for the FAS known to suffice for T-cell recognition, it has been difficult to determine for which human tumors the foreign antigen pool is sufficiently large to elicit T-cell reactivity. To provide such a benchmarked analysis of the foreign antigen loads in human tumors, we selected three viral references that have been shown to provide sufficient genomically foreign T-cell antigens to yield clinically meaningful responses. First, the E6 and E7 oncoproteins from human papilloma virus (HPV) which are expressed

in cervical cancer, head and neck cancer, and ano-genital cancers, but also in premalignant neoplastic lesions. Vaccination of patients with premalignant vulvar intraepithelial neoplasia with pools of overlapping HPV16 E6 and E7 peptides has previously been demonstrated to lead to regression of lesions in more than 80% of patients²⁷. Thus, clinically meaningful T-cell reactivity against epitopes within the E6 and E7 oncoproteins can apparently be induced in a large fraction of patients. Second, the EBV LMP-1 and LMP-2 oncogenes which are causally involved in EBV-induced transformation of a number of tissue types and the expression of these genes continues after initial cellular transformation. Evidence for clinically relevant immunogenicity of LMP-1/LMP-2 is provided by the observation that T-cell products generated against these two proteins mediate regression of a large fraction of EBV-positive relapsed or refractory Hodgkin and non-Hodgkin lymphomas (OR of 62%, CR of 52%; (Bollard et al., 2014; Heslop et al., 2010). Third, as a non-cancer related viral reference, numerous CTL responses have been identified in patients infected with HIV³⁴. Furthermore, reports on both humans and non-human primates have shown a correlation between HIV-specific CD8⁺ T cells and control of viremia^{29,30}, and evidence for CTL-mediated control of HIV-1 is also provided by the association of slow disease progression with certain HLA alleles and presence of conserved T-cell epitopes^{34–37}.

To use these viral gene sets as benchmarks, their predicted foreign antigen loads are required to scale similarly to viral immunogenicity as those from human genes. As gene expression influences cell surface antigen abundance^{38–40} and as such likely T-cell recognition probability, structural differences between viral and human gene expression would complicate the intercomparison of foreign antigen loads. We compared the expression levels of the benchmark viral oncogenes and human protein-coding genes in tissue samples maximally similar to those in the studies we used to define our benchmarks. To the best of our knowledge, RNA sequencing data of VIN lesions used for the HPV E6/E7 benchmark is not available, so we analyzed HPV-positive high-grade cervical intraepithelial neoplasia (CIN) lesions. Here we found the average expression of HPV E6/7 to be slightly lower (0.77-fold difference in medians, $p = 0.83$; Figure S2.3A) than human genes, which is unlikely to be of discernible effect on cell surface antigen abundance as RNA expression levels vary over 4 orders of magnitude (Fig S3A). In an analogous analysis of EBV⁺ diffuse large B-cell lymphomas, average expression of EBV *LMP-1* and *LMP-2* was found to be ~7.7-fold lower (0.13-fold difference in medians, $p = 0.2$; Figure S2.3B) than human genes. The trend towards somewhat lower expression of the viral oncogenes, renders these oncogenes a reasonable and possibly slightly conservative benchmark for the foreign antigen load known to suffice for the induction of clinically relevant T cell reactivity.

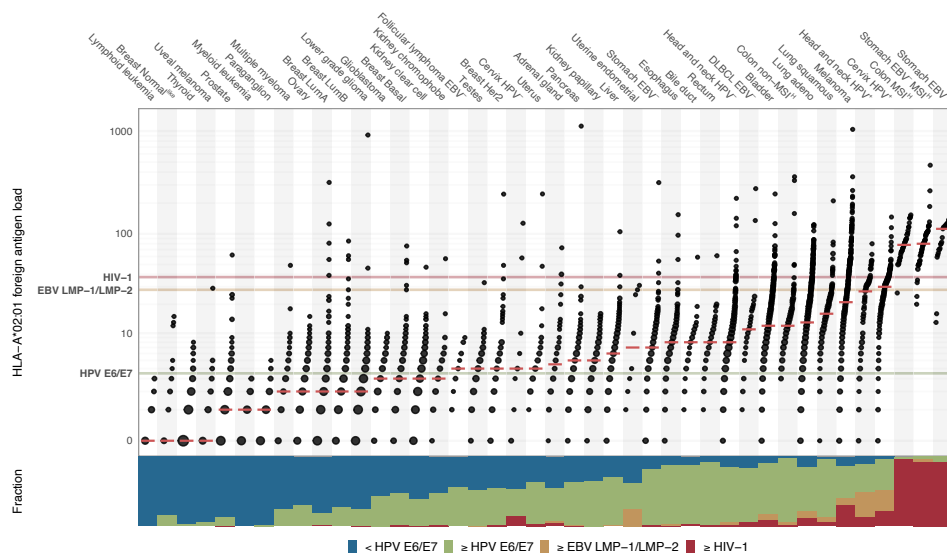
Benchmarking the foreign antigen space of human malignancies against human pathogens

To generate a comprehensive estimate of immunological foreignness across cancer types, we combined the predicted HLA*A:02:01-neo-antigen loads resulting from somatic DNA damage, with those

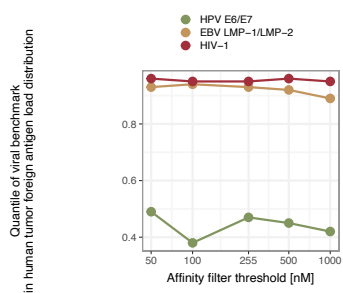
of non-germline sequences in the rearranged B-cell receptor in case of B-cell malignancies and of expressed viral genes for virus-positive cancers. For the latter class of genes, viral contributions were restricted to genes expressed within the top three quartiles of the human transcriptome. As we previously found neoantigen depletion to be incomplete at best (Chapter), we analyzed HLA-A*02:01 neo-antigen predictions as a proxy for all class I neo-antigens for all patients, regardless of their HLA repertoires.

Comparing the foreign antigen loads of human tumors to those of our viral benchmarks, we found large overall consistency with observed clinical successes of cancer immunotherapy obtained so far⁴¹, while simultaneously also highlighting tumor types worthy of further exploration in terms of immunotherapeutic targeting of foreign antigens. 16 out of the 42 evaluated tumor types had a median foreign antigen load below the lowest of the 3 benchmarks, the HPV E6/E7 oncogenes, which were enriched for hematological and neurological tumors (Figure 2.3A). Our current analysis provides no evidence for the average foreign antigen repertoire of these tumors to be sufficiently sized for clinical actionability. The upper side of the foreignness spectrum is composed of cancers with foreign antigen score that are larger than the HPV E6/E7 benchmark, comprising 53% of all assayed samples. Here, 4/26 tumor types and 9% of analyzed samples had higher median foreign antigen loads than the EBV LMP-1/LMP-2 proteins. In view of the profound clinical activity of T cell products directed against these antigens, these data provide strong evidence that the repertoire of neo-antigens present in these tumors should commonly suffice to allow strong T cell reactivity. Notably, virus-positive tumors, including HPV⁺-tumors, all displayed median foreign antigen loads higher than that of the EBV LMP-1/LMP-2 oncoprotein-benchmark. Foreignness benchmarks were robust to variation of the MHC-affinity thresholds of the prediction pipeline, keeping the other pipeline parameters constant (Figure 2.3B).

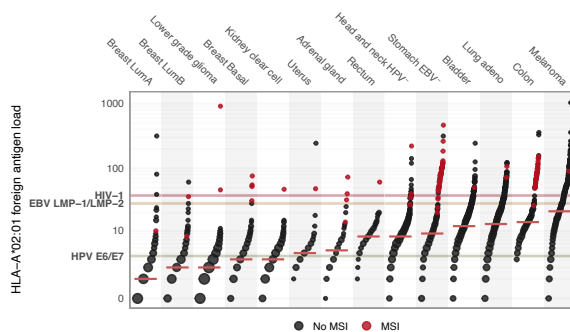
A



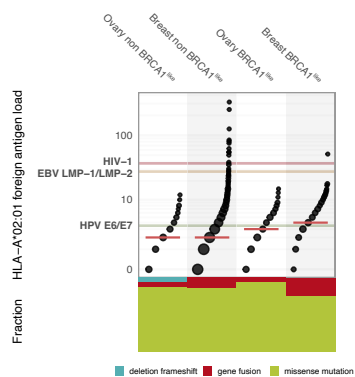
B



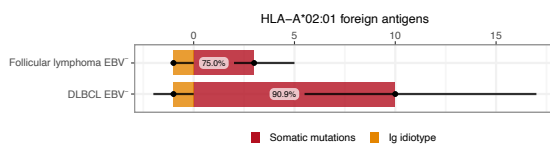
C



D



E



F

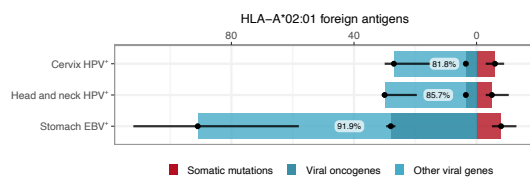


Figure 2.3: Pan-cancer overview of immunological foreignness

A (Top) Pan-cancer foreign antigen loads contrasted to the foreign antigen load of three viral gene sets known to allow the formation of protective T-cell responses. Tumoral foreign antigen loads are composed of neoantigens generated through focal DNA damage, virus-derived antigens (virus-positive tumors only) and immunoglobulin idiotype-derived antigens (B-cell malignancies only). Fusion gene contributions are omitted as these could not be evaluated for all patients and their contribution was modest (see Figure 2.2). Dot size reflects the number of samples on a coordinate, red lines indicate median foreign antigen load per tumor type. (Bottom) Distributions of tumoral foreignness relative to indicated viral benchmark gene sets.

B Robustness of the benchmarking of tumor foreignness to the peptide MHC-affinity threshold used in neo-antigen predictions. Quantiles of the three viral benchmarks within the distributions of tumoral foreign antigen loads are depicted.

C Distribution of microsatellite instable tumors across indicated tumor types. Only those TCGA-tumor types in which at least one MSI^H-case was detected (shown as red dots) are depicted.

D Foreign antigen loads and compositions of breast and ovarian cancers classified by presence or absence of BRCA1-like signature characteristics in copy number profiles (Schouten et al., 2015).

E Foreign antigen contributions of focal DNA damage and immunoglobulin idiotype sequences in B-cell lymphomas. Black dots denote medians, black bars span the 25% and 75% percentiles. Percentage labels denote the contribution of viral genes to the overall foreign antigen load.

F Foreign antigen contributions of focal DNA damage and either expressed HPV genes in HPV⁺ head and neck and cervical tumors, or expressed EBV genes in EBV⁺ stomach adenocarcinomas. Black dots denote medians, black bars span the 25% and 75% percentiles. Percentage labels denote the contribution of viral genes to the overall foreign antigen load.

Mismatch repair deficiency enlarges tumor foreignness

Consistent with the high response rates to checkpoint blockade^{15,16}, mismatch-repair deficient colorectal and non-colorectal cancers are predicted to have very high neoantigen loads, deriving from somatic mutations in repetitive regions of the genome, termed microsatellites (Figure 2.3C). We assessed the microsatellite instability (MSI)-status of all TCGA samples in our cohort with a classifier trained on mutation type frequencies⁴² and found MSI in 12/29 assayed tumor types (1.8% of assayed samples) and that the foreign antigen loads of 91% of MSI^H-tumors are higher than that of our upper benchmark, the full HIV genome (Figure 2.3C). This indicates that MSI-tests can be used to directly identify highly foreign tumors across tumor types. Although mismatch-repair proficient colon cancers had lower neoantigen loads than their mismatch-repair deficient counterparts, virtually all tested exceeded our lower HPV E6/E7 benchmark. This suggests that the low response rate to PD-1 blockade that has been observed in this tumor type¹⁵ cannot solely be explained by a lack of immunological foreignness, suggesting an enrichment of immune evasion mechanisms there.

Defects in homology-mediated DNA repair augment foreignness

Defects in homology-mediated DNA repair (e.g. by loss of function of the *BRCA1/2* genes) have been associated with large-scale genomic rearrangements such as inter- and intrachromosomal gene fusion events in breast and ovarian cancers^{43–45}. These rearrangements can yield protein-coding tran-

scripts with fusions either occurring in-frame, leading to novel junctional sequence at the boundary of the two fused exons, or out-of-frame, generating complete stretches of novel coding sequences. To assess whether defects in homologous recombination DNA repair are associated to the complete foreign antigen load, we stratified patients of the TCGA breast and ovarian cancer cohorts according to their 'BRCA-like' status, as assessed with a 'BRCA1-like' classifier trained to recognize DNA copy number aberration profiles characteristic for *BRCA1*-mutants⁴⁶. BRCA1-like breast cancer was significantly more foreign than non-BRCA1-like breast cancer (1.75-fold difference in medians, $p = 5 \times 10^{-10}$), reaching median foreign antigen loads equal to our lower benchmark HPV E6/E7 (Figure 2.3D, top), partially due to a higher presence of gene fusions (Figure 2.3D, bottom). Similarly, BRCA1-like patients were significantly more foreign than non-BRCA1-likes in the ovarian carcinoma cohort (1.5-fold difference in medians, $p = 6 \times 10^{-3}$), but neither group's median reached the lower HPV E6/7 benchmark. Consistent with our observations, an association has been found between *BRCA1/2*-mutation status and elevated neo-antigen loads in high-grade serous ovarian cancer⁴⁷.

Composition of foreign antigens informs on novel potential therapeutic targets

As vaccination strategies in cancers with viral etiology have primarily focused on the targeting of viral antigens, we wished to explore how the foreign antigen loads of viral oncogenes in these tumors compares to that formed as a consequence of somatic mutations. Follicular B-cell lymphomas, despite being very low in exonic non-synonymous mutations (median: 30, range: [9, 79]), get the primary contribution to their foreignness from neoantigens originating from somatic mutations (Figure 2.3E). Due to their low numbers, idiotype-derived antigens (median: 1, range: [0, 4]) are less attractive therapeutic targets, compared to the more abundant DNA damage-derived neoantigens. Similarly, the neoantigens generated by somatic mutations in the highly mutated diffuse large B-cell lymphomas (DLBCL) easily surpass those of idiotypic sequences in these tumors, and it will be interesting to establish whether long-term responses of DLBCL following CD19 CAR T-cell therapy⁴⁸ are in some cases accompanied by induction of neoantigen-specific T-cell reactivity. On the contrary, for HPV⁺ cervix and head & neck and EBV⁺-stomach cancers, the majority of the foreign antigen repertoire is made up by antigens of viral origin (Figure 2.3F). even when restricting the viral contribution to the LMP-1/LMP-2 oncogenes

Gene fusions have been suggested as another potentially important source of foreign antigens^{49,50}. We detected 11426 gene fusion events in 3359 out of 6049 investigated tumor samples across 33 malignancies. Likely due to their low prevalence, the overall contribution of fusion genes to foreignness was modest (6.8% of total foreignness; 95% CI: [6.3%, 7.2%]). This suggests low prioritization of immunotherapeutic solely and specifically targeting fusion gene antigens. Notable exceptions to this include prostate, uterine and some breast cancers (Luminal B and Basal subtypes) for which fusion-derived neoantigens increased the foreignness to surpass that of our lower HPV E6/E7 benchmark

(Figure S2.4A).

As we found NMD to have a substantial impact on neoantigen load (Figure 2.2A), especially for antigens derived from indels, a potential immunostimulatory measure would be to inhibit NMD to augment the targetable antigenome^{51,52}. By including epitopes from NMD-flagged transcripts in our foreignness estimates, we find that an average of 10% (95% CI: [10%, 11%]) of HLA-A*02:01 neoantigens could be gained (Figure S2.4B). As expected, the highest gains were in indel-rich, mismatch-repair deficient tumor types (colon MSI^H: 32.2%, stomach MSI^H: 31.2%, Figure S2.4B). The evaluated kidney cancers, through their elevated indel loads, additionally showed gains of 16% (95% CI: [14%, 18%]) in foreignness when accounting for NMD-blockade, averaged over all three evaluated subtypes. We found up to 14% of clear cell kidney tumors to have potentially clinically impactful increases in foreignness, as they met or surpassed the HPV E6/E7 benchmark after NMD-blockade (Figure S2.4B).

Accounting for intratumoral heterogeneity consistently lowers foreign antigen space

Clonal antigens that are presented by all tumor cells in a lesion can reasonably be expected to have a larger contribution to tumor regression than subclonal antigens, and the observed inverse relationship between tumor heterogeneity and immunotherapy outcome provides indirect support for the superior value of clonal antigens as T-cell targets⁵³. We aimed to address how our foreignness estimates would be affected when accounting for intratumoral heterogeneity.

To account for tumor heterogeneity in our antigenome estimates, we devised two ways of incorporating cellularity-estimates in our predicted neoantigen loads. In a first, conservative approach, we fully restricted predicted antigenomes to those peptides that are derived from clonal mutations (i.e. with an inferred cellularity $\geq .95$ ^{53,54}). In this, we assumed the contribution of viral genes to be shared by all cells in HPV⁺ and EBV⁺ cancers due to their role in cellular transformation. We observed an average 51.6% (95% CI: [50.5%, 52.7%]) decrease in these heterogeneity-aware foreign antigen loads as compared to the heterogeneity-unaware foreign antigen loads presented earlier (Figure 2.4A), decreasing the number of tumor types with median foreign antigen loads equal or exceeding our lower benchmark HPV E6/E7 from 12/18 to 8/18 (Figure S2.5A). The above analysis ignores that subclonal mutations can be immunogenic and could contribute to tumor rejection, limited to the subclones harboring them. Indeed, T-cell recognition of subclonal mutations has been reported⁵⁵. This inspires an approach in which neoantigens are weighted by the cellularity of their associated somatic mutations. We found cellularity-weighted estimates of the foreign antigen loads to be 23.5% lower (95% CI: [22.7%, 24.3%]), when compared to heterogeneity-unaware foreign antigen loads (Figure 2.4A,B), more mildly decreasing the number of tumor types with median foreign antigen loads equal or exceeding our lower benchmark HPV E6/E7 from 12/18 to 10/18. Interestingly, the decrease in foreign antigen load when accounting for ITH is not constant across tumor types. Whereas

the antigenic mutanomes of all breast cancer subtypes and melanoma were relatively rich in nearly clonal mutations and hence experienced only mild decreases in foreign antigen loads (median decreases ranging between 8.2% and 13.6%), thyroid cancer on the other side of the spectrum underwent a median decrease of 42.2% (Figure 2.1A). Read coverage depth in next-generation sequencing is known to influence the sensitivity of variant callers to (subclonal) mutations⁵⁶. However, the variation in effect of cellularity weighting was not explained by differences in median read coverage of variant loci (Figure S2.5B).

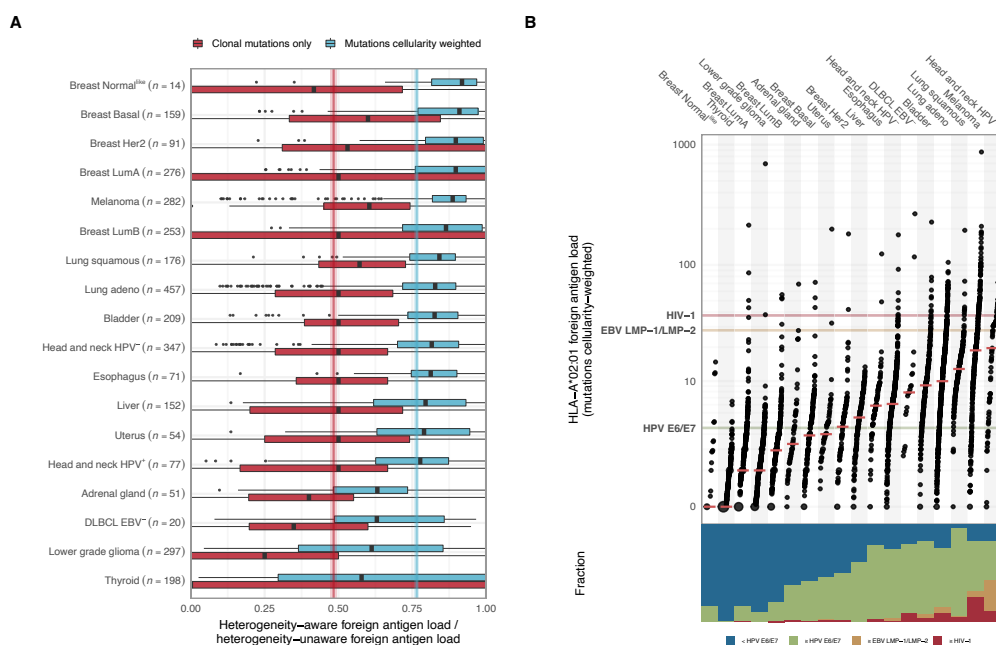


Figure 2.4: Effect of tumor heterogeneity on foreign antigen loads

A. Distributions of the ratio between heterogeneity-aware and -unaware somatic mutation-derived foreign antigen estimates. Colored lines reflect the average decrease in foreign antigen loads when either restricting the computation of the foreign antigen loads to clonal mutations (52%, 95% CI: [50%, 53%, red), or when weighting mutations according to their cellularity (23%, 95% CI: [23%, 24%], blue line).

B. Pan-cancer overview of foreign antigen loads in which neo-antigens contributed by focal DNA damage types are weighted by their estimated cellularity and combined with viral and idiotype antigen contributions.

Discussion

We devised an unbiased method to estimate the immunological foreignness resulting from different forms of DNA damage and oncogenic viral insertions in human cancers, and then benchmarked these

results against the foreignness of human pathogens for which clinically relevant T-cell recognition has convincingly been demonstrated. The main conclusion of this work is that a substantial fraction of human cancers expresses a sufficient number of neo-antigens to allow T cell recognition.

By establishing the precision of our epitope prediction pipeline, and by using the same prediction pipeline on foreign genomes known to contain T cell targets, we removed concerns around the unknown false positive rates that have complicated the interpretation of earlier antigenome landscape efforts. One striking outcome of this benchmarking is that a substantial set of human cancers for which no forms of immunotherapy have been approved to date (e.g. MSS colon, bile duct, MSS endometrial, pancreatic, adrenocortical, and uterine cancers) do contain a larger number of predicted neo-antigens than the HPV E6 and E7 oncogenes that have successfully been used as therapeutic targets in premalignant disease. One interpretation of these data is that defects at other stages in the cancer immunity cycle/cancer immunogram^{57,58} are interfering with T cell-mediated tumor control in these tumor types. For tumor types, such as MSS colon cancer, that do commonly encode a number of neo-antigens that lies above that of the HPV E6 and E7 oncogenes, it will be of interest to determine whether inefficient T cell priming, or the immunosuppressive molecules such as TGF- β , may for instance explain the poor responsiveness to PD-1/ PD-L1 blockade.

In future efforts, the accuracy of the estimates of foreignness that we provide may be further improved in several ways. First, the contribution of multi-basepair indels and gene fusion events is potentially underestimated as sequencing and the employed variant calling methodologies to detect these events were less robust than for SNVs^{59–61}. Effect size of such optimization would be expected to be most significant for tumor types that are characterized by chromosomal instability and by relatively low degrees of focal DNA damage, such as *BRCA1/2* deficient cancers. A second point of improvement will be to include the weakly compensatory interaction between MHC-binding affinity and transcription levels in cell surface antigen abundance, as shown by mass-spectrometry of HLA-eluted peptides^{38,40,62,63}. However, as gene expression levels between the viral HPV and EBV benchmarks and the tumor-associated transcripts differ only minimally, we expect the effect of inclusion of a weighted measure of RNA expression/MHC affinity to be small. Analysis of the contribution of subclonal versus clonal antigens to the total pool of predicted neo-antigens revealed that for tumor types such as breast and melanoma accounting for intratumoral heterogeneity only minimally alters the size of this pool whereas for thyroid carcinoma more than half of patients completely lost their foreign antigen repertoire. In particular for this tumor type it will be important to understand to what extent subclonal T cell antigens can contribute to tumor control, for instance by development of mouse models with a fixed level of clonal and increasing levels of subclonal neo-antigens.

Based on the data from the Melief and Bollard/Rooney groups it is apparent that induction of T cell reactivity against HPV E6/ E7 and EBV LMP1/LMP2 frequently results in regression of premalignant HPV⁺ lesions and EBV⁺ lymphomas, respectively^{27,28}. An important and incompletely resolved question is whether these therapy-induced T cell responses are by themselves sufficient, or whether the observed clinical responses are at least partially due to epitope spreading. In case of the HPV anti-

gens, such a dominant role of novel T cell responses as a consequence of epitope spreading appears less likely, as the premalignant lesions that respond to vaccination²⁷ are unlikely to have accumulated substantial numbers of somatic mutations⁶⁴. Nevertheless, a better understanding of epitope spreading in therapy-induced tumor control would be of value.

In earlier work by Turajlic and colleagues, the neo-antigen yield rate of frameshifting indels was found to be 9-fold higher than that of non-synonymous SNVs³², whereas in our analyses these DNA mutation types differed only minimally in this respect. The discrepancy between these data likely stems from the omission of a model for NMD and the use of a model for thymic negative selection of neo-antigen-reactive T cells that is strict on SNV-derived peptides but not applied to peptides originating from indels, whereas SNV and indel-derived epitopes in our work were both passed through the similarity-to-self filter.

We have provided a pan-cancer overview of immunological foreignness that predicts a potential value of cancer immunotherapies that goes substantially beyond the tumor types for which T cell checkpoint currently forms an approved strategy, and that includes common cancers such as MSS CRC. Analysis of the mechanisms that hold back T cell reactivity in these tumors and strategies to circumvent these mechanisms will be of major interest.

Bibliography

1. Brahmer, J., Reckamp, K.L., Baas, P., Crino, L., Eberhardt, W.E., Poddubskaya, E., Antonia, S., Pluzanski, A., Vokes, E.E., Holgado, E., et al. (2015). Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* 373, 123–135.
2. Ferris, R.L., Blumenschein, Jr., G., Fayette, J., Guigay, J., Colevas, A.D., Licitra, L., Harrington, K., Kasper, S., Vokes, E.E., Even, C., et al. (2016). Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* 375, 1856–1867.
3. Topalian, S.L., Sznol, M., McDermott, D.F., Kluger, H.M., Carvajal, R.D., Sharfman, W.H., Brahmer, J.R., Lawrence, D.P., Atkins, M.B., Powderly, J.D., et al. (2014). Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J Clin Oncol* 32, 1020–1030.
4. Gettinger, S.N., Horn, L., Gandhi, L., Spigel, D.R., Antonia, S.J., Rizvi, N.A., Powderly, J.D., Heist, R.S., Carvajal, R.D., Jackman, D.M., et al. (2015). Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol* 33, 2004–2012.
5. Motzer, R.J., Escudier, B., McDermott, D.F., George, S., Hammers, H.J., Srinivas, S., Tykodi, S.S., Sosman, J.A., Procopio, G., Plimack, E.R., et al. (2015). Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 373, 1803–1813.

-
6. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128.
 7. Robert, C., Long, G.V., Brady, B., Dutriaux, C., Maio, M., Mortier, L., Hassel, J.C., Rutkowski, P., McNeil, C., Kalinka-Warzocha, E., et al. (2015). Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* 372, 320–330.
 8. Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 348, 62–68.
 9. Draper, L.M., Kwong, M.L., Gros, A., Stevanovic, S., Tran, E., Kerkar, S., Raffeld, M., Rosenberg, S.A., and Hinrichs, C.S. (2015). Targeting of HPV-16+ Epithelial Cancer Cells by TCR Gene Engineered T Cells Directed against E6. *Clin Cancer Res* 21, 4431–4439.
 10. Stevanovic, S., Draper, L.M., Langan, M.M., Campbell, T.E., Kwong, M.L., Wunderlich, J.R., Dudley, M.E., Yang, J.C., Sherry, R.M., Kammula, U.S., et al. (2015). Complete regression of metastatic cervical cancer after treatment with human papillomavirus-targeted tumor-infiltrating T cells. *J Clin Oncol* 33, 1543–1550.
 11. Tume, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 515, 568–571.
 12. Dudley, M.E., Gross, C.A., Somerville, R.P., Hong, Y., Schaub, N.P., Rosati, S.F., White, D.E., Nathan, D., Restifo, N.P., Steinberg, S.M., et al. (2013). Randomized selection design trial evaluating CD8+-enriched versus unselected tumor-infiltrating lymphocytes for adoptive cell therapy for patients with melanoma. *J Clin Oncol* 31, 2152–2159.
 13. Tran, E., Ahmadzadeh, M., Lu, Y.C., Gros, A., Turcotte, S., Robbins, P.F., Gartner, J.J., Zheng, Z., Li, Y.F., Ray, S., et al. (2015). Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* 350, 1387–1390.
 14. van Rooij, N., van Buuren, M.M., Philips, D., Velds, A., Toebe, M., Heemskerk, B., van Dijk, L.J., Behjati, S., Hilkman, H., El Atmioui, D., et al. (2013). Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol* 31, e439–42.
 15. Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 372, 2509–2520.
 16. Le, D.T., Durham, J.N., Smith, K.N., Wang, H., Bartlett, B.R., Aulakh, L.K., Lu, S., Kemberling, H., Wilt, C., Luber, B.S., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357, 409–413.

17. Powles, T., Eder, J.P., Fine, G.D., Braiteh, F.S., Loriot, Y., Cruz, C., Bellmunt, J., Burris, H.A., Petrylak, D.P., Teng, S.L., et al. (2014). MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515, 558–562.
18. Wolchok, J.D., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.-J., Cowey, C.L., Lao, C.D., Wagstaff, J., Schadendorf, D., Ferrucci, P.F., et al. (2017). Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* 377, 1345–1356.
19. Tran, E., Robbins, P.F., Lu, Y.C., Prickett, T.D., Gartner, J.J., Jia, L., Pasetto, A., Zheng, Z., Ray, S., Groh, E.M., et al. (2016). T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J Med* 375, 2255–2262.
20. Tran, E., Turcotte, S., Gros, A., Robbins, P.F., Lu, Y.C., Dudley, M.E., Wunderlich, J.R., Somerville, R.P., Hogan, K., Hinrichs, C.S., et al. (2014). Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* 344, 641–645.
21. Verdegaal, E.M., de Miranda, N.F., Visser, M., Harryvan, T., van Buuren, M.M., Andersen, R.S., Hadrup, S.R., van der Minne, C.E., Schotte, R., Spits, H., et al. (2016). Neoantigen landscape dynamics during human melanoma-T cell interactions. *Nature* 536, 91–95.
22. Segal, N.H., Parsons, D.W., Peggs, K.S., Velculescu, V., Kinzler, K.W., Vogelstein, B., and Allison, J.P. (2008). Epitope landscape in breast and colorectal cancer. *Cancer Res* 68, 889–892.
23. Rajasagi, M., Shukla, S.A., Fritsch, E.F., Keskin, D.B., DeLuca, D., Carmona, E., Zhang, W., Sougnez, C., Cibulskis, K., Sidney, J., et al. (2014). Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 124, 453–462.
24. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61.
25. Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* 165, 35–44.
26. Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211.
27. Kenter, G.G., Welters, M.J., Valentijn, A.R., Lowik, M.J., Berends-van der Meer, D.M., Vloon, A.P., Essahsah, F., Fatherson, L.M., Offringa, R., Drijfhout, J.W., et al. (2009). Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. *N Engl J Med* 361, 1838–1847.
28. Bollard, C.M., Gottschalk, S., Torrano, V., Diouf, O., Ku, S., Hazrat, Y., Carrum, G., Ramos, C., Fayad, L., Shpall, E.J., et al. (2014). Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. *J. Clin. Oncol.* 32, 798–808.

-
29. Koup, R.A., Safrit, J.T., Cao, Y., Andrews, C.A., McLeod, G., Borkowsky, W., Farthing, C., and Ho, D.D. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J. Virol.* 68, 4650–4655. D - NLM: PMC236393 EDAT- 1994/07/01 MHDA- 1994/07/01 00:01 CRDT- 1994/07/01 00:00 PST - ppublish.
30. Kuroda, M.J., Schmitz, J.E., Charini, W.A., Nickerson, C.E., Lifton, M.A., Lord, C.I., Forman, M.A., and Letvin, N.L. (1999). Emergence of CTL coincides with clearance of virus during primary simian immunodeficiency virus infection in rhesus monkeys. *J. Immunol. Baltim. Md* 1950 162, 5127–5133.
31. Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 8, 33.
32. Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., et al. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* 18, 1009–1021.
33. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
34. Allen, T.M., Yu, X.G., Kalife, E.T., Reytor, L.L., Lichterfeld, M., John, M., Cheng, M., Allgaier, R.L., Mui, S., Frahm, N., et al. (2005). De novo generation of escape variant-specific CD8+ T-cell responses following cytotoxic T-lymphocyte escape in chronic human immunodeficiency virus type 1 infection. *J Virol* 79, 12952–12960.
35. Kaslow, R.A., Carrington, M., Apple, R., Park, L., Munoz, A., Saah, A.J., Goedert, J.J., Winkler, C., O'Brien, S.J., Rinaldo, C., et al. (1996). Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* 2, 405–411.
36. Migueles, S.A., Sabbaghian, M.S., Shupert, W.L., Bettinotti, M.P., Marincola, F.M., Martino, L., Hallahan, C.W., Selig, S.M., Schwartz, D., Sullivan, J., et al. (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U A* 97, 2709–2714.
37. O'Connor, D.H., Mothe, B.R., Weinfurter, J.T., Fuenger, S., Rehrauer, W.M., Jing, P., Rudersdorf, R.R., Liebl, M.E., Krebs, K., Vasquez, J., et al. (2003). Major Histocompatibility Complex Class I Alleles Associated with Slow Simian Immunodeficiency Virus Disease Progression Bind Epitopes Recognized by Dominant Acute-Phase Cytotoxic-T-Lymphocyte Responses. *J. Virol.* 77, 9029–9040.
38. Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G.L., Eisenhaure, T.M., et al. (2017). Mass Spectrometry Profiling of HLA-Associated

- Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* 46, 315–326.
39. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., and Mann, M. (2015). Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics* 14, 658–673.
 40. Juncker, A.S., Larsen, M.V., Weinhold, N., Nielsen, M., Brunak, S., and Lund, O. (2009). Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. *PLoS One* 4, e7448.
 41. Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* 377, 2500–2501.
 42. Huang, M.N., McPherson, J.R., Cutcutache, I., Teh, B.T., Tan, P., and Rozen, S.G. (2015). MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci Rep* 5, 13321.
 43. Konishi, H., Mohseni, M., Tamaki, A., Garay, J.P., Croessmann, S., Karnan, S., Ota, A., Wong, H.Y., Konishi, Y., Karakas, B., et al. (2011). Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc Natl Acad Sci U S A* 108, 17773–17778.
 44. Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54.
 45. Venkitaraman, A.R. (2014). Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science* 343, 1470–1475.
 46. Schouten, P.C., Grigoriadis, A., Kuilman, T., Mirza, H., Watkins, J.A., Cooke, S.A., van Dyk, E., Severson, T.M., Rueda, O.M., Hoogstraat, M., et al. (2015). Robust BRCA1-like classification of copy number profiles of samples repeated across different datasets and platforms. *Mol Oncol* 9, 1274–1286.
 47. Strickland, K.C., Howitt, B.E., Shukla, S.A., Rodig, S., Ritterhouse, L.L., Liu, J.F., Garber, J.E., Chowdhury, D., Wu, C.J., D'Andrea, A.D., et al. (2016). Association and prognostic significance of BRCA1/2-mutation status with neoantigen load, number of tumor-infiltrating lymphocytes and expression of PD-1/PD-L1 in high grade serous ovarian cancer. *Oncotarget* 7, 13587–13598.
 48. Neelapu, S.S., Locke, F.L., Bartlett, N.L., Lekakis, L.J., Miklos, D.B., Jacobson, C.A., Braunschweig, I., Oluwole, O.O., Siddiqi, T., Lin, Y., et al. (2017). Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med* 377, 2531–2544.
 49. Chang, T.C., Carter, R.A., Li, Y., Li, Y., Wang, H., Edmonson, M.N., Chen, X., Arnold, P., Geiger, T.L., Wu, G., et al. (2017). The neoepitope landscape in pediatric cancers. *Genome Med* 9, 78.

-
50. Capietto, A.-H., Hoshyar, R., and Delamarre, L. (2022). Sources of Cancer Neoantigens beyond Single-Nucleotide Variants. *Int. J. Mol. Sci.* *23*, 10131.
 51. El-Bchiri, J., Guilloux, A., Dartigues, P., Loire, E., Mercier, D., Buhard, O., Sobhani, I., de la Grange, P., Auboeuf, D., Praz, F., et al. (2008). Nonsense-mediated mRNA decay impacts MSI-driven carcinogenesis and anti-tumor immunity in colorectal cancers. *PLoS One* *3*, e2583.
 52. Pastor, F., Kolonias, D., Giangrande, P.H., and Gilboa, E. (2010). Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. *Nature* *465*, 227–230.
 53. McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* *351*, 1463–1469.
 54. Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* *152*, 714–726.
 55. Anagnostou, V., Smith, K.N., Forde, P.M., Niknafs, N., Bhattacharya, R., White, J., Zhang, T., Adleff, V., Phallen, J., Wali, N., et al. (2017). Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov* *7*, 264–276.
 56. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* *31*, 213–219.
 57. Blank, C.U., Haanen, J.B., Ribas, A., and Schumacher, T.N. (2016). CANCER IMMUNOLOGY. The “cancer immunogram.” *Science* *352*, 658–660.
 58. Chen, D.S.S., and Mellman, I. (2013). Oncology meets immunology: the cancer-immunity cycle. *Immunity* *39*, 1–
 59. Cornish, A., and Guda, C. (2015). A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference. *Biomed Res Int* *2015*, 456479.
 60. Hwang, S., Kim, E., Lee, I., and Marcotte, E.M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep* *5*, 17875.
 61. Liu, S., Tsai, W.H., Ding, Y., Chen, R., Fang, Z., Huo, Z., Kim, S., Ma, T., Chang, T.Y., Priedigkeit, N.M., et al. (2016). Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res* *44*, e47.
 62. Fortier, M.H., Caron, E., Hardy, M.P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P. (2008). The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* *205*, 595–6
 63. Garcia Alvarez, H.M., Koşaloğlu-Yalçın, Z., Peters, B., and Nielsen, M. (2022). The role of antigen expression in shaping the repertoire of HLA presented ligands. *iScience* *25*, 104975.

64. Rusan, M., Li, Y.Y., and Hammerman, P.S. (2015). Genomic landscape of human papillomavirus-associated cancers. *Clin Cancer Res* 21, 2009–2019.
65. Fritsch, E.F., Rajasagi, M., Ott, P.A., Brusic, V., Hacohen, N., and Wu, C.J. (2014). HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res* 2, 522–529.
66. Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., and Verhaak, R.G. (2015). The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* 34, 4845–4854. 10.1038/onc.2014.406.
67. Torres-Garcia, W., Zheng, S., Sivachenko, A., Vegesna, R., Wang, Q., Yao, R., Berger, M.F., Weinstein, J.N., Getz, G., and Verhaak, R.G. (2014). PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics* 30, 2224–2226. 10.1093/bioinformatics/btu169.

Methods

Datasets

HIV validation, TCGA mutation calls and RNA expression was acquired and processed as described in **Chapter 4**.

ICGC data

CLLE-ES (chronic lymphocytic leukemia) and MALY-DE (malignant lymphoma) projects were downloaded from ICGC release 20. Somatic mutations were annotated with RNA expression levels from sample-matched expression data, matched on Ensembl gene identifiers.

CGCI data

Diffuse large B-cell lymphoma (DLBCL) RNA sequencing data was obtained from the NHL-DLBCL project, provided by the Cancer Genome Characterization Initiative.

CoMMpass data

Multiple myeloma DNA and RNA sequencing data was obtained from the CoMMpass website (<https://research.themmr.org/>), provided by the Multiple Myeloma Research Foundation, in January 2016. Somatic variants were annotated with sample-matched gene expression levels from the provided gene expression matrix, matched on Ensembl gene identifiers.

Genotype-Tissue Expression Project (GTEx) data

Healthy spleen and EBV-transformed lymphocyte datasets were obtained from the GTEx Portal (db-GaP project phs000424) and used to quantify background and positive control EBV transcript expression levels. The raw data were processed as described under 'Viral gene expression'.

B-cell lymphoma idiotype data

Idiotype sequences obtained from Diffuse Large B-cell lymphoma (DLBCL) and Follicular lymphoma (FL) patients were provided by the Veelken lab at the Leiden University Medical Center, The Netherlands. In selected DLBCL cases where two heavy or light chains were identified, both chains were included. The FL dataset was restricted to heavy chain sequencing data.

Gene fusion calls

Gene fusion calls for the majority of TCGA donors were provided by the Verhaak lab at The University of Texas MD Anderson Cancer Center, USA¹, identified using the structural variant caller PRADA². Of note, PRADA only detects sense-sense fusions, a limitation that to the best of our knowledge applies to all fusion gene callers. To determine the amino acid sequence of gene fusions, the canonical nucleotide sequences of both fusion partners were retrieved from Ensembl release 64 as this version of Ensembl was used for the fusion variants in the PRADA analysis. Genomic breakpoint coordinates were converted to cDNA coordinates and mRNA features (e.g. UTRs) at the breakpoint locations were annotated. Fusions that were deemed unlikely to result in a translatable amino acid sequence (e.g. UTR-UTR fusions) were discarded. The nucleotide sequences of fusion partners in the remaining fusion events were joined and translated into amino acid sequences. As out-of-frame fusions result in the loss of the canonical stop codon, the nucleotide sequence after the breakpoint was scanned until the first occurrence of an in-frame stop codon and the resulting sequence was translated to protein. The resulting amino acid sequence was subsequently used in the neo-antigen prediction pipeline described below. As the PRADA-analysis was performed on RNASeq data, NMD-targeting of detected fusion transcripts can at best have been incomplete, but we cannot exclude partial degradation of these fusion variant calls. For this reason, we did not apply the NMD-filter to fusion variant calls.

Genome- and patient-level annotation

MMR status of tumor samples, antigen presentation capability and sensitivity to T-cell attack, PAM50 subtyping of breast cancer samples, gene essentiality were performed as described in **Chapter 4**.

Peptide-level annotation

Similarity to self-repertoire of peptides was performed as described in **Chapter 4**.

Somatic variant annotation

Variant effect prediction, variant oncogenicity and essentiality, variant cellularity, variant expression, neoantigen prediction, validation of the neoantigen prediction pipeline were performed as described in **Chapter 4**.

Variant annotation with the likelihood of having been caused by any mutational process

Methods to extract the mutational signature of mutational processes compute the total burden caused by a mutational process and the mutational signatures of these processes for a large collection of genomes. They do not compute probabilities for single mutations to have been originated by a particular process, as one would require to compare the neo-antigen yielding capabilities of different mutational processes. We devised a methodology to do this, starting from mutational profiles and patient specific signature loadings.

We want to compute the likelihood of the n -th signature to have caused a somatic variant v of the i th type/letter in the alphabet of all 192 point mutations with the analyzed strand selected such that the reference allele is a pyrimidine and flanked by all possible nucleotide combinations (e.g. 'ApC', denoting a C to T mutation for which the 5' and 3' flanking variants are adenosine and guanine, respectively, employing the notation in Alexandrov et al. (2013)³³). This measure should factor in the relative contribution of signature n , f^n , to the genome g in which the variant was observed. The absolute amount of mutational signature explained mutations of type i in genome g contributed by the n -th signature, $N_{n,g}^i$, equals the forward probability of the n -th signature to cause a variant of the i -th letter multiplied by the presence of the n -th signature in the genome g : $N_{n,g}^i = p_{n,g}^i e_g^n$ where e_g^n represents the total contribution of signature n to genome g . From this, the probability for mutations of type i to have been caused by the n -th signature in that same genome g must be $\frac{N_{n,g}^i}{\sum_{n'} N_{n',g}^i}$. Somatic variants were assigned to the different mutational processes proportional to the likelihood of having been caused by these probabilities, after which somatic variant and neo-antigen loads could be tallied cognizant of partial class membership as described in **Chapter 4** 'Somatic variant and neo-epitope load tallying'.

Neoantigen load prediction

Computing the total foreign antigen load/foreignness

Foreign antigen load was computed as the sum of the neo-antigen load and viral antigen load (where applicable) on a sample-specific basis. For DLBCL and FL, idiotypic-derived antigen loads were added to this by adding the medians of idiotypic-derived foreign antigen loads found in non-matched DLBCL and FL patients, respectively.

Viral contribution to foreign antigen load

We noticed substantial differences in the expression levels of EBV genes and human genes in EBV⁺ stomach cancer. In order not to overestimate the viral contribution to the foreign antigen load, we filtered out genes that had lower expression values than the first quartile of the human distribution on a per sample basis. Application of this rule had a relatively minor impact on the antigen contributions of HPV genes in HPV⁺ cervical and head and neck tumors, with all viral genes being expressed at a higher level than the first quartile of the human protein-encoding transcriptome in the majority of tumors (68% and 59%, respectively) were unaffected due to all viral genes being expressed to a higher extent than the first quartile of human genes. Affected tumors lost an average of 42% (95% CI: [35%, 48%]) and 48% (95% CI: [36%, 61%]) of viral antigens due to low expression, respectively (Figure S2.3B). In contrast, all EBV⁺ stomach tumors were strongly affected, as 64% (95% CI: [58%, 70%]) of viral antigens derive from genes that are expressed at a lower level than the first quartile of human protein-encoding transcriptome (Figure S2.3C).

Computation of heterogeneity aware neo-antigen loads

We computed two kinds of heterogeneity aware foreign antigen loads. First, the ‘clonals only’ measure, in which only neo-antigens derived from clonal mutations are included. Second, the ‘cellularity-weighted’ measure, in which neo-antigens are summed, weighted by the maximum likelihood cellularity estimates of their corresponding mutations. For both of these measures, we assumed the contribution of viral genes to be shared by all cells in HPV⁺ and EBV⁺ cancers due to their role in cellular transformation. In addition, idiotypic antigen contributions were weighted as one.

Statistical analysis

Antigen prediction quality assessment, somatic variant and neo-epitope load tallying was performed in as described in **Chapter 4**.

Neoantigen yield rate computation

To compare the neoantigen generating propensities of different somatic variant classes, we analyze their yield rates, defined for a class c as the number of neoantigens resulting from c divided by the mutation load contributed by c . When comparing the yield rates of different mutation classes c , they are computed for the aggregate of variants and neoantigens from multiple donors in a group of genomes G :

$$r_{c,G} = \frac{\sum_{g \in G} N_{g,c}^p}{\sum_{g \in G} N_{g,c}^v}$$

which can be interpreted as the probability for a mutation of class c to yield a neoantigen. Invoking the binomial properties of the r for values of r in the range $[0, 1]$, we compute the 95% confidence interval of this estimate as $r_{c,G} \pm 1.96 \sqrt{\frac{1}{N_{g,c}^n} r_{c,G} (1 - r_{c,G})}$.

Data and software availability

Custom software was distributed over R and Perl packages as described in **Chapter 4**.

Supplemental Items

Table 1 can be retrieved from <https://doi.org/10.17632/mf39n7s2b9.1>.



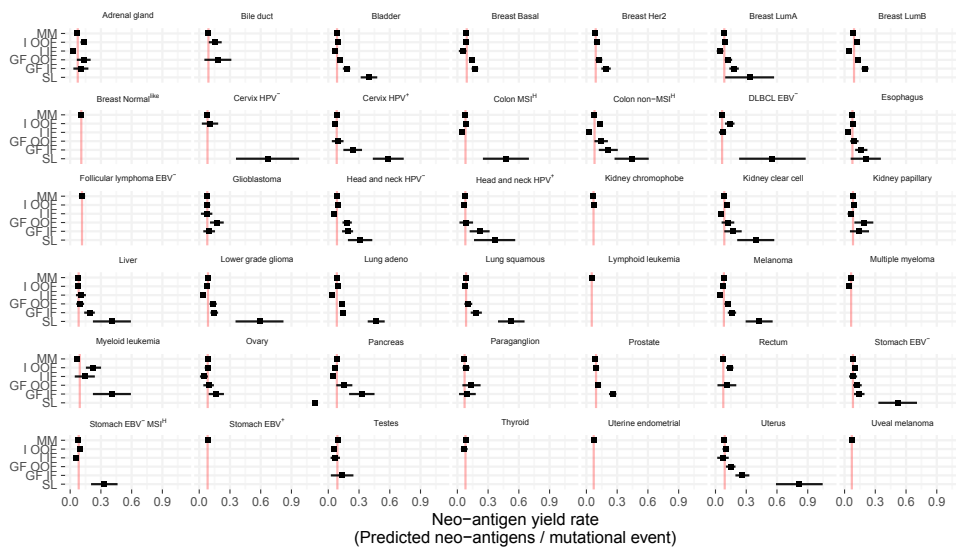
Figure S2.1: Technical considerations in assessing foreignness

A Distance in base pairs between fusion gene breakpoints and focal DNA damage (SNVs and indels of max. 200 base pairs long).

B The pairwise Spearman correlation in predicted neoantigen loads between different settings of the neoantigen prediction pipeline across all included patients. Rows are identically ordered as columns. Gene expression filtering settings, rather than the choice of the HLA allele to use in neoantigen prediction, determine most of the variation. This indicates the choice of HLA allele is rather unimportant.

A

HLA-A*02:01 - PR=1.9 STS Prot=0.5 Exp=0-NMD



B

HLA-A*02:01 - PR=1.9 STS Prot=0.5 Exp=0-NMD

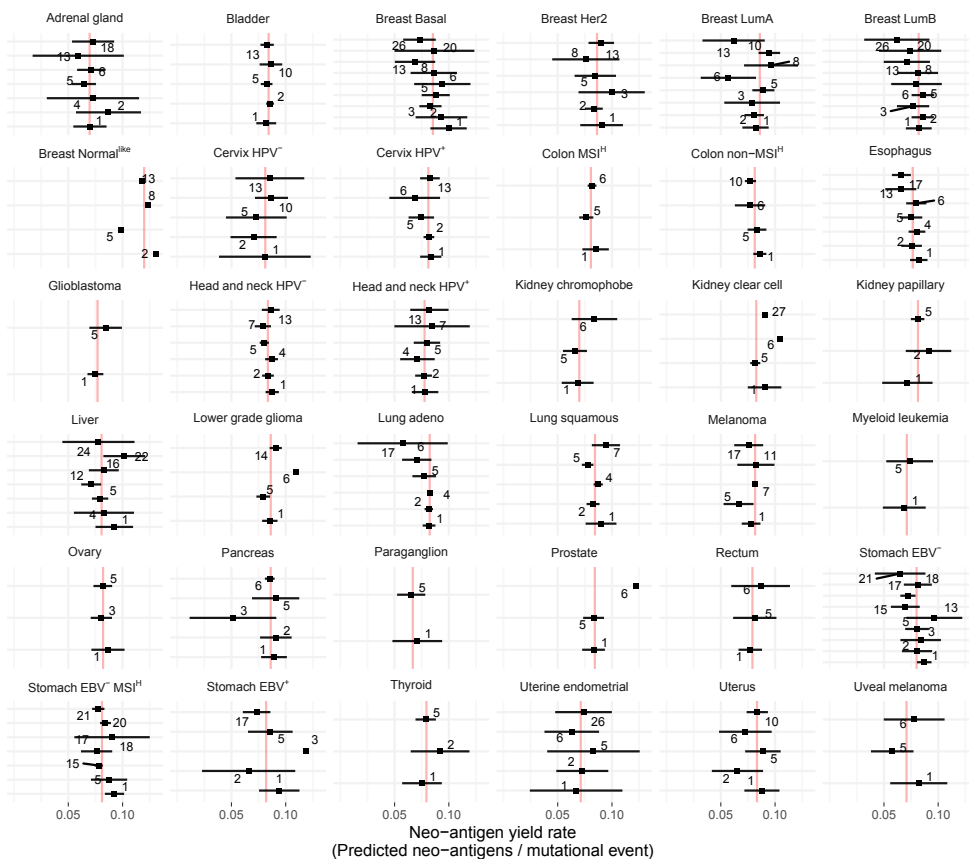


Figure S2.2: Relationship between mutational processes and foreign antigen loads across cancers.

A HLA-A*02:01 neo-antigen yield rates of different mutational events are highly similar across tumor types. Tumor-specific average yield rates are indicated by red lines.

B As in **A**, Comparing neo-antigen yield rates of the indicated mutational signatures within tumor types. No signatures are observed to be recurrently deviant from the other mutational signatures with respect to neo-antigen yield rates.

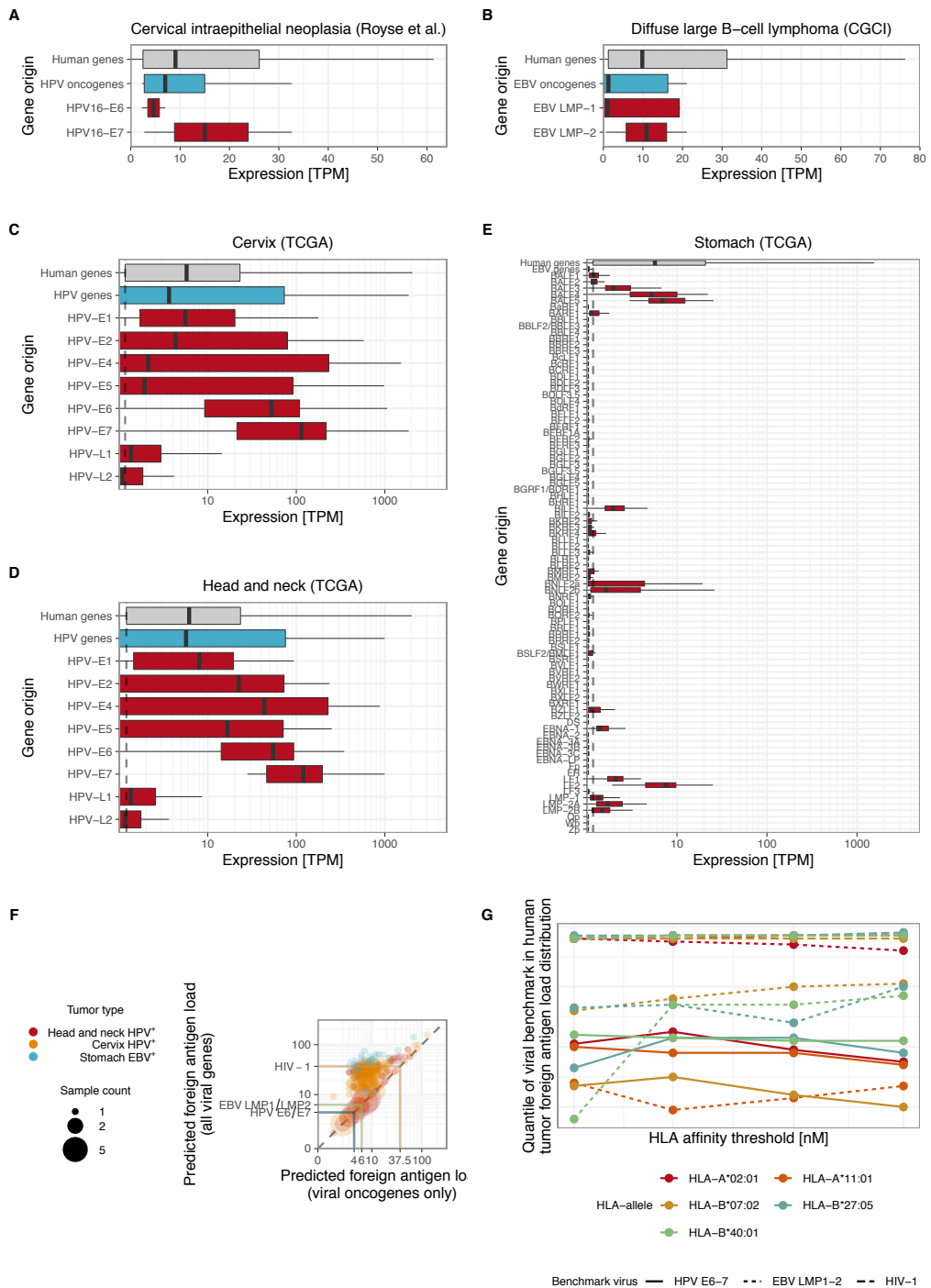


Figure S2.3: Validation of comparisons of viral antigens and antigens derived from somatic mutations

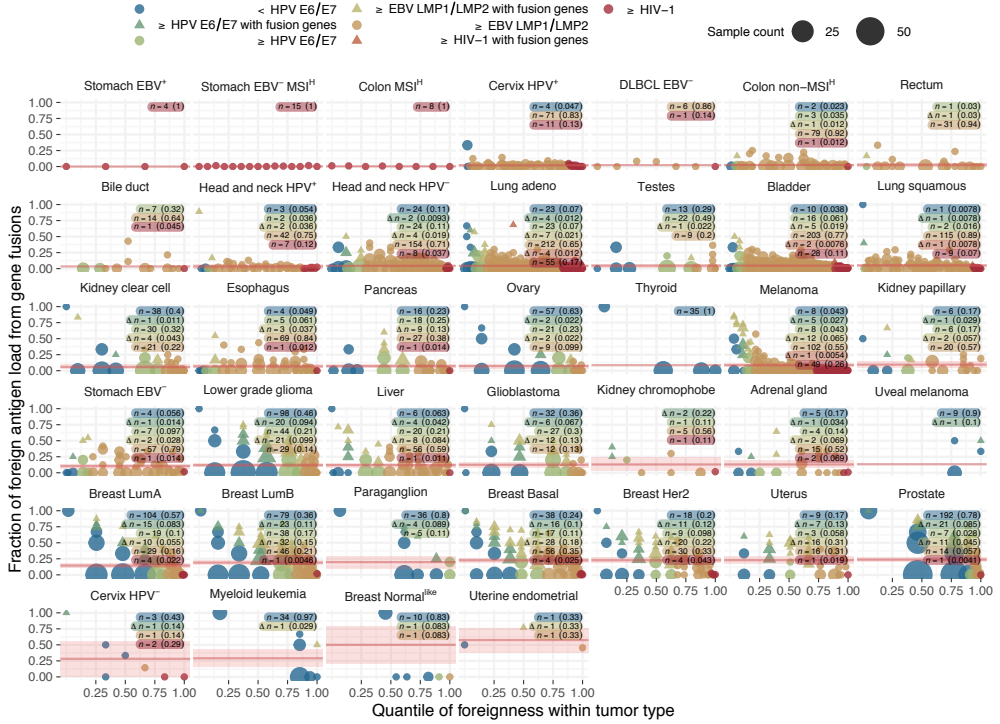
A (Left) Distributions of RNA expression levels of human genes and HPV-associated *E6* and *E7* oncogenes in cervical intraepithelial neoplasia. (Right) Same as left panel, but depicting expression of human genes and EBV-derived oncogenes in diffuse large B-cell lymphoma. Blue boxes represent expression levels of viral oncogenes.

B (Top) Distributions of RNA expression levels of human genes and HPV genes in cervical cancers. (Bottom) As in top panel, but for HPV⁺ head and neck cancers. Blue boxes represent expression of all viral genes.

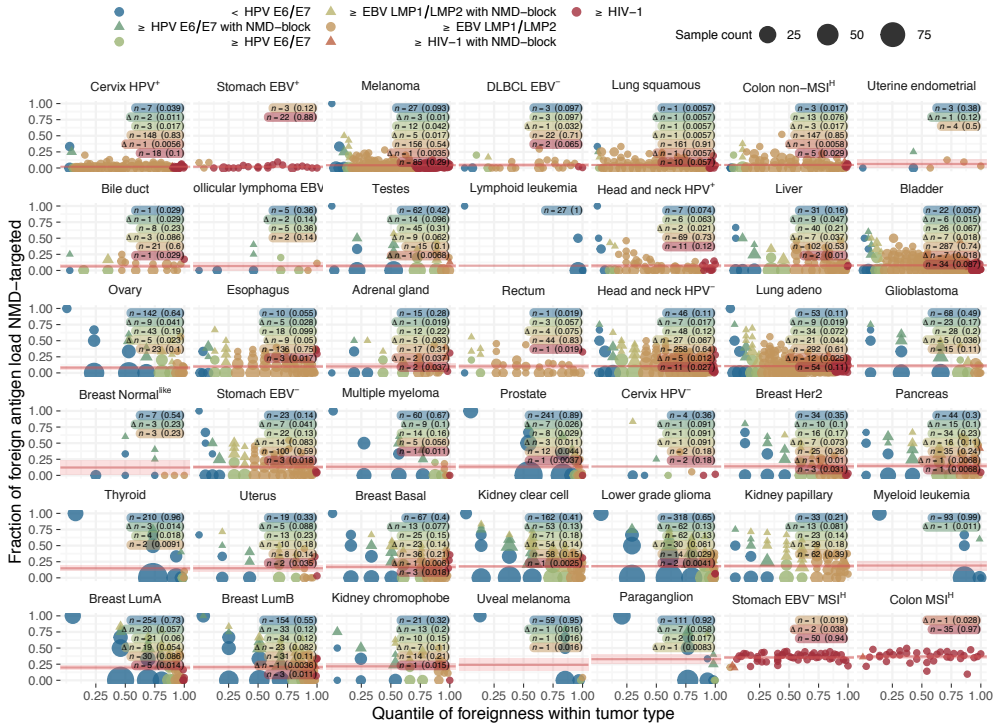
C RNA expression levels of human genes and EBV genes in EBV⁺ stomach tumors. Dashed line indicates lower quartile expression level of human genes. The blue box represents expression across all EBV genes.

D Relationship between foreign antigen loads of virus-positive tumors when either only taking viral oncogenes (*E6* and *E7* for HPV, *LMP-1* and *LMP-2* for EBV) into account, or when taking all viral genes into account for which expression greater than the first quartile of the human genes is observed in a sample-specific manner. Dot size reflects the number of samples on a coordinate; color denotes tumor type.

A

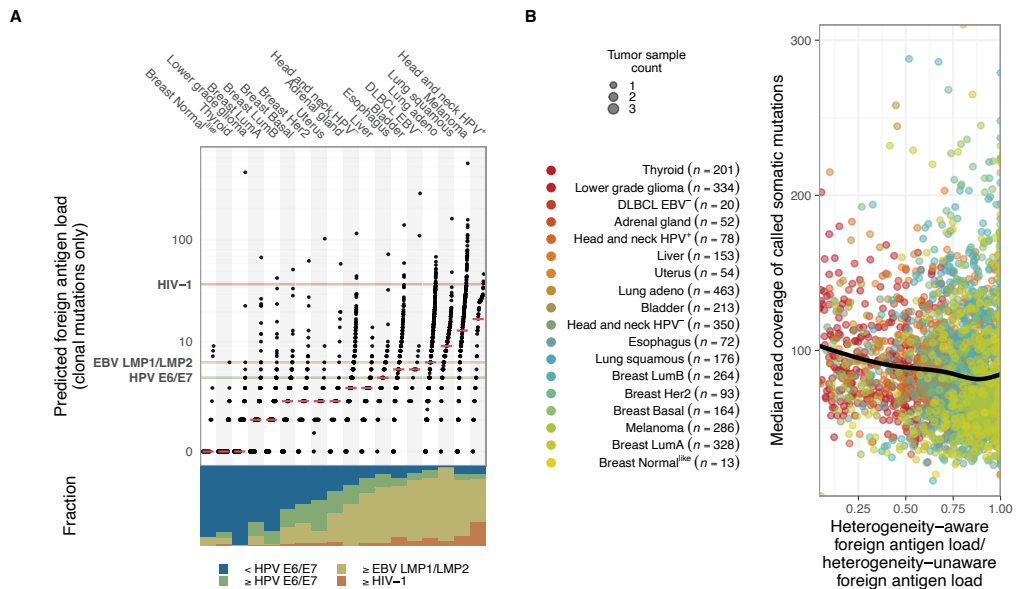


B



A Relationship between the foreign antigen loads excluding fusion genes (quantile within tumor type, horizontal axis) and the fractional contribution of fusion-derived antigens to the neo-antigen load. Dot size reflects the number of samples on a coordinate. Only tumor samples for which the presence of gene fusion events was evaluated are incorporated. Samples are colored according to their foreignness compared to the three viral benchmarks. Tumors moving up to a higher level of foreignness, as compared to the viral benchmarks, when including fusion-derived antigens are denoted by triangles rather than circles. Legends indicate sample size per color group, in which black parenthesized numbers reflect the fraction of samples in each color group per tumor type. Mean fractional contributions and associated 95% confidence intervals of fusion genes to total foreignness are indicated by red lines and surrounding red transparent boxes, respectively. Rank quantiles shown on the horizontal axis are computed over the foreign antigen load excluding fusion genes, assigning tied samples identical maximal rank quantiles. Fusions genes are shown to infrequently contribute a sizeable fraction of foreign antigen load, especially for tumors that are relatively rich in focal mutation-derived foreign antigens.

B As in **A**, but showing the fractional contribution of NMD-targeted neo-antigens to the tumoral foreign antigen loads.



A Pan-cancer overview of foreignness in which neo-antigens contributed by focal DNA damage types are restricted to clonal mutations and combined with viral and idiosyncratic antigen contributions, with the latter two antigen sources assumed to be clonal.

B Absence of relationship between sample DNA sequencing depth and the degree to which the foreign antigen load is reduced in heterogeneity aware estimates. Dot size reflects the number of samples at a coordinate. The black line represents a LOESS local smoothing regression.

IMMUNE INDUCTION STRATEGIES IN METASTATIC TRIPLE NEGATIVE
BREAST CANCER TO ENHANCE THE SENSITIVITY TO PD-1 BLOCKADE:
THE TONIC-TRIAL

Leonie Voorwerk^{*}, Maarten Slagter^{*}, Hugo M. Horlings, Karolina Sikorska, Koen K. van de Vijver, Michiel de Maaker, Iris Nederlof, Roelof J.C. Kluin, Sarah Warren, SuFey Ong, Terry G. Wiersma, Nicola S. Russell, Ferry Lalezari, Philip C. Schouten, Noor A.M. Bakker, Steven L.C. Ketelaars, Dennis Peters, Charlotte A.H. Lange, Erik van Werkhoven, Harm van Tinteren, Ingrid A. M. Mandjes, Inge Kemper, Suzanne Onderwater, Myriam Chalabi, Sofie Wilgenhof, John B.A.G. Haanen, Roberto Salgado, Karin E. de Visser, Gabe S. Sonke, Lodewyk F.A. Wessels, Sabine C. Linn, Ton N. Schumacher, Christian U. Blank and Marleen Kok

^{*} These authors contributed equally

Nature Medicine, 2019.

DOI: 10.1038/s41591-019-0432-4

Abstract

The efficacy of programmed cell death protein 1 (PD-1) blockade in metastatic triple-negative breast cancer (TNBC) is low¹⁻⁵, highlighting a need for strategies that render the tumor microenvironment more sensitive to PD-1 blockade. Preclinical research has suggested immunomodulatory proper-

ties for chemotherapy and irradiation⁶⁻¹³. In the first stage of this adaptive, non-comparative phase 2 trial, 67 patients with metastatic TNBC were randomized to nivolumab (1) without induction or with 2-week low-dose induction, or with (2) irradiation (3 × 8 Gy), (3) cyclophosphamide, (4) cisplatin or (5) doxorubicin, all followed by nivolumab. In the overall cohort, the objective response rate (ORR; iRECIST¹⁴) was 20%. The majority of responses were observed in the cisplatin (ORR 23%) and doxorubicin (ORR 35%) cohorts. After doxorubicin and cisplatin induction, we detected an upregulation of immune-related genes involved in PD-1–PD-L1 (programmed death ligand 1) and T cell cytotoxicity pathways. This was further supported by enrichment among upregulated genes related to inflammation, JAK–STAT and TNF- α signaling after doxorubicin. Together, the clinical and translational data of this study indicate that short-term doxorubicin and cisplatin may induce a more favorable tumor microenvironment and increase the likelihood of response to PD-1 blockade in TNBC. These data warrant confirmation in TNBC and exploration of induction treatments prior to PD-1 blockade in other cancer types.

Main text

Triple-negative breast cancer (TNBC), characterized by estrogen receptor, progesterone receptor and HER2 negativity, comprises 10–20% of all breast cancers¹⁵. In patients with metastatic disease, tumors rapidly become resistant to chemotherapy, resulting in a median overall survival of only 8–13 months^{16,17}. Although durable responses to PD-1 and programmed death-ligand 1 (PD-1/PD-L1) blockade have been observed in TNBC, the fraction of patients with metastatic TNBC that benefit from PD-1/PD-L1 blockade is low, with response rates around 5%^{1,4}. Response rates seem to increase to 19–23% upon selection of patients with PD-L1-positive tumor microenvironments (TMEs)^{2,18}. However, the majority of patients with TNBC do not benefit from PD-1/ PD-L1 blockade, highlighting the need for strategies that can alter the immune-suppressive TME and increase sensitivity to PD-1/ PD-L1 blockade.

Table 1: Baseline characteristics of the intention-to-treat population. Clinical baseline characteristics of all allocated patients. PD-L1 immunohistochemistry was performed using the DAKO 22C3 clone.

Total population (n = 70)	
Median age, years (range)	51 (29-70)
WHO performance status, n (%)	
0	41 (59%)
1	29 (41%)
Germline BRCA1/2, n (%)	
Mutation	6 (9%)
Wildtype	50 (71%)
Unknown	14 (20%)
Location of metastasis, n (%)	

Lymph node only	6 (9%)
Visceral metastasis	50 (71%)
Other metastasis	14 (20%)
Number of previous therapies for metastatic disease, n (%)	
0	17 (24%)
1	34 (49%)
2-3	19 (27%)
Previous neoadjuvant or adjuvant therapy, n (%)	
Yes	59 (84%)
Previous chemotherapy exposure, n (%)	
Taxane	64 (91%)
Anthracycline	60 (86%)
Platinum	42 (60%)
Capecitabine	34 (49%)
Disease Free Interval (DFI), n (%)	
De novo metastatic disease	9 (13%)
DFI \leq 12 months	23 (33%)
DFI $>$ 12 months	38 (54%)
LDH level, n (%)	
\leq ULN	39 (56%)
\leq 2x ULN	31 (44%)
PD-L1 expression on tumor cells, n (%)	
Not available	5 (7%)
\geq 1% on tumor cells	44 (63%)
\geq 5% on tumor cells	23 (33%)
PD-L1 expression on immune cells, n (%)	
Not available	5 (7%)
\geq 1% on immune cells	60 (86%)
\geq 5% on immune cells	47 (67%)

Preclinical and clinical studies have shown that low-dose chemotherapy or irradiation may be utilized to stimulate anticancer immune responses. For example, irradiation has been shown to induce type I interferons via the stimulator of interferon genes (STING) pathway and consequently enhance T cell priming^{6,7}. Some studies have demonstrated that cyclophosphamide can deplete regulatory T cells and could restore effector functions of T cells and natural killer cells⁸. In addition, cisplatin has been shown to upregulate major histocompatibility complex class I expression and directly stimulate T cell function^{9,10}. Finally, doxorubicin has been associated with myeloid-derived suppressor cell (MDSC) depletion¹¹, an increase in the level of type I interferons¹² and induction of immunogenic cell death¹³.

Here, we present a phase 2 trial in which we dissect the immunomodulatory effects of hypofractionated irradiation and low-dose cyclophosphamide, cisplatin and doxorubicin in patients with TNBC,

with the hypothesis that these treatments may be utilized as priming strategies to improve the efficacy of PD-1/PD-L1 blockade. This multi-cohort TONIC trial evaluates the efficacy of nivolumab after short-term induction with low-dose chemotherapy, irradiation or no induction. A ‘pick-the-winner’ strategy, taking into account clinical responses and translational findings, was used with a Simon’s two-stage design¹⁹ to decide which cohorts would be expanded.

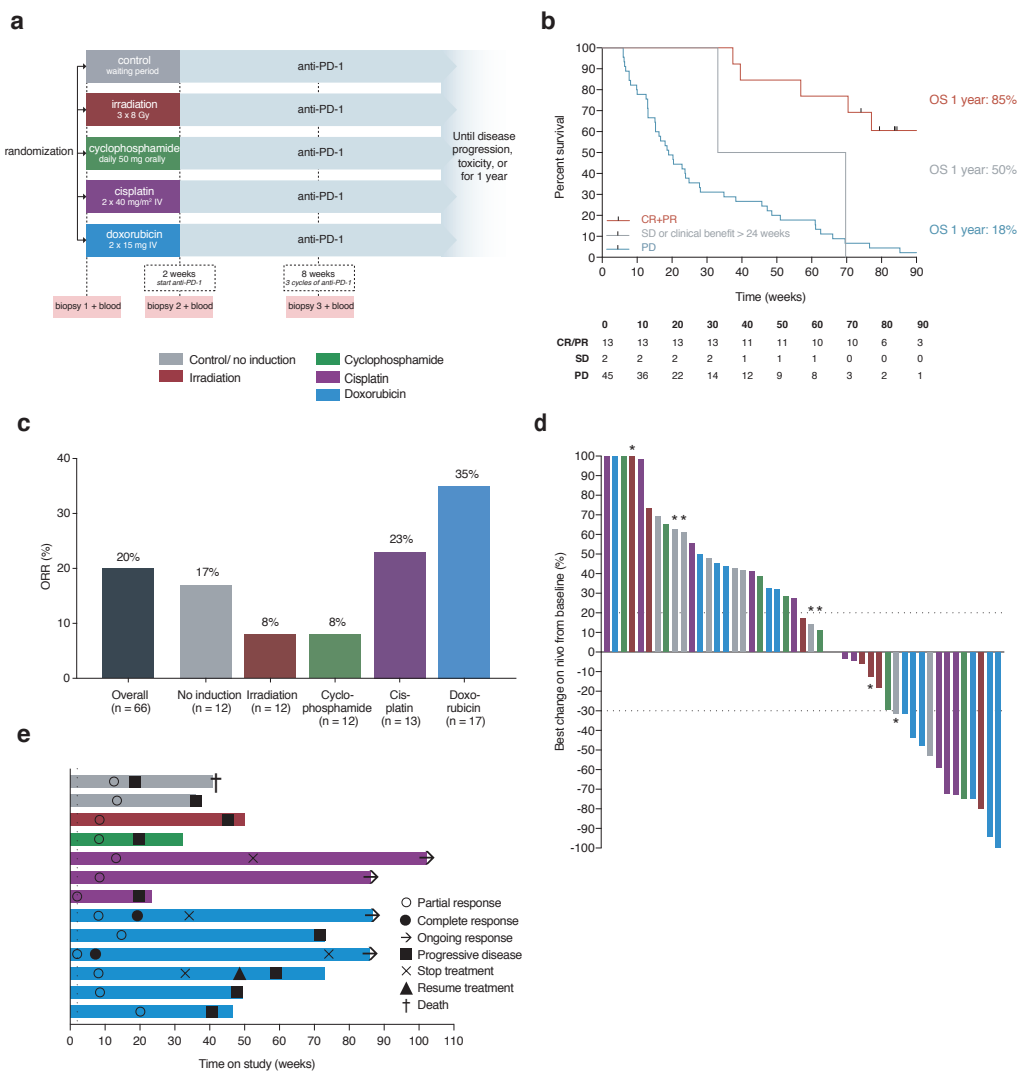


Figure 3.1: Anti-tumor activity of nivolumab after immune induction in the per protocol population.
a Design of the TONIC trial. Patients were randomized to 1 of 4 cohorts with induction treatments or no in-

duction, all followed by nivolumab (3 mg per kg every 2 weeks). Biopsies and blood samples were taken at baseline (biopsy one), on post-induction treatment (biopsy two) and on nivolumab (after three cycles of nivolumab; biopsy three). i.v., intravenous.

b Overall survival (OS) by response. Kaplan–Meier curves of overall survival by best overall response were calculated. All 67 patients of the per protocol population were included, but 7 patients were deceased within 6 weeks after nivolumab initiation, and their data are not displayed (that is, a landmark was used at 6 weeks). The stable disease (SD) group includes a patient with stable disease, as defined by RECIST, for 26 weeks and a patient with non-evaluable disease but clinical benefit for 26 weeks. PD, progressive disease.

c ORR per cohort as the percentage of total patients per cohort (iRECIST, investigator determined). ORR comprises all PRs and CRs.

d Waterfall plot. Best radiological response of target lesions during nivolumab treatment compared to baseline. Eleven patients with clinical evidence of disease progression did not have a follow-up CT scan after nivolumab initiation, and nine patients had non-measurable disease. Depicted is the largest change in the sum of target lesions, in comparison to baseline or the post-induction CT scan (changes compared to the post-induction scan are indicated by asterisks; $n = 7$). Bar colors reflect the induction treatment shown in a. The y axis was cut-off at 100% for illustration purposes. Dotted black lines indicate the response as described by RECIST1.1.

e Swimmers plot. Duration of response of patients with PR or CR according to iRECIST. Progressive disease was assessed according to iRECIST; the first date of progressive disease is depicted in case of confirmation on a subsequent CT scan. Only two patients had a PR after induction treatment, with one prolongation after nivolumab treatment. One patient with a microsatellite instable tumor, pretreated with cisplatin, ended treatment after 1 year and has had an ongoing remission for 102 weeks. One patient with a CR stopped treatment after 17 nivolumab cycles due to a grade 2 pneumonitis and has had an ongoing CR for 86 weeks; another patient with a CR stopped treatment due to a grade 2 gastritis after 38 cycles of nivolumab and has had an ongoing CR for 86 weeks. The vertical dotted line marks the 2-week induction period.

In the TONIC trial (NCT02499367), patients were randomized to one of four different induction treatments, consisting of irradiation to a single lesion, low-dose cyclophosphamide, cisplatin or doxorubicin, or a 2-week waiting period (Figure 3.1a). Biopsies from metastatic lesions were taken at baseline (biopsy one), after induction (biopsy two) and after three cycles of nivolumab (biopsy three). Seventy patients were randomized between September 2015 and October 2017. Accrual continued until a minimum of ten patients who received at least one cycle of nivolumab and from whom we could acquire high-quality paired biopsies were included for each cohort, resulting in a slightly uneven number of patients across cohorts (Figure S3.2). At data cut-off, the median follow-up was 19.9 months. Characteristics were as expected for advanced TNBC (Table 1) and balanced between cohorts, with a relatively high proportion of patients in the doxorubicin and control cohorts receiving their first-line treatment in this trial (Supplementary Table S1). Sixty-six patients were available for efficacy analysis (Supplementary Table S2). All patients had received previous chemotherapy in the (neo)-adjuvant and/or the metastatic setting. Patients with de novo stage IV disease ($n = 8$ out of 66) were pretreated with palliative chemotherapy before entering the TONIC trial.

Nivolumab after induction was not associated with any previously unreported toxicity. Induction treatment-related adverse events (AEs) of any grade occurred in 19 patients (28%, with 3% grade 3) and immune-related AEs of grades 3–5 occurred in 13 patients (19%; Supplementary Tables S3 and S4). Two patients with evidence of progression died on study.

Median progression-free survival (PFS) for all patients was 1.9 months (Supplementary Table S5). We observed an objective response rate (ORR) to nivolumab of 20% (13 out of 66 patients; iRECIST¹⁴), with two complete responses (CRs; 3%) and 11 partial responses (PRs; 17%) (Supplementary Table S5). The median duration of response according to iRECIST was 9 months (95% CI: 4.7 not reached). At data lock, four patients were still on study: one patient was still receiving nivolumab with an ongoing response, and three patients were in remission after stopping nivolumab.

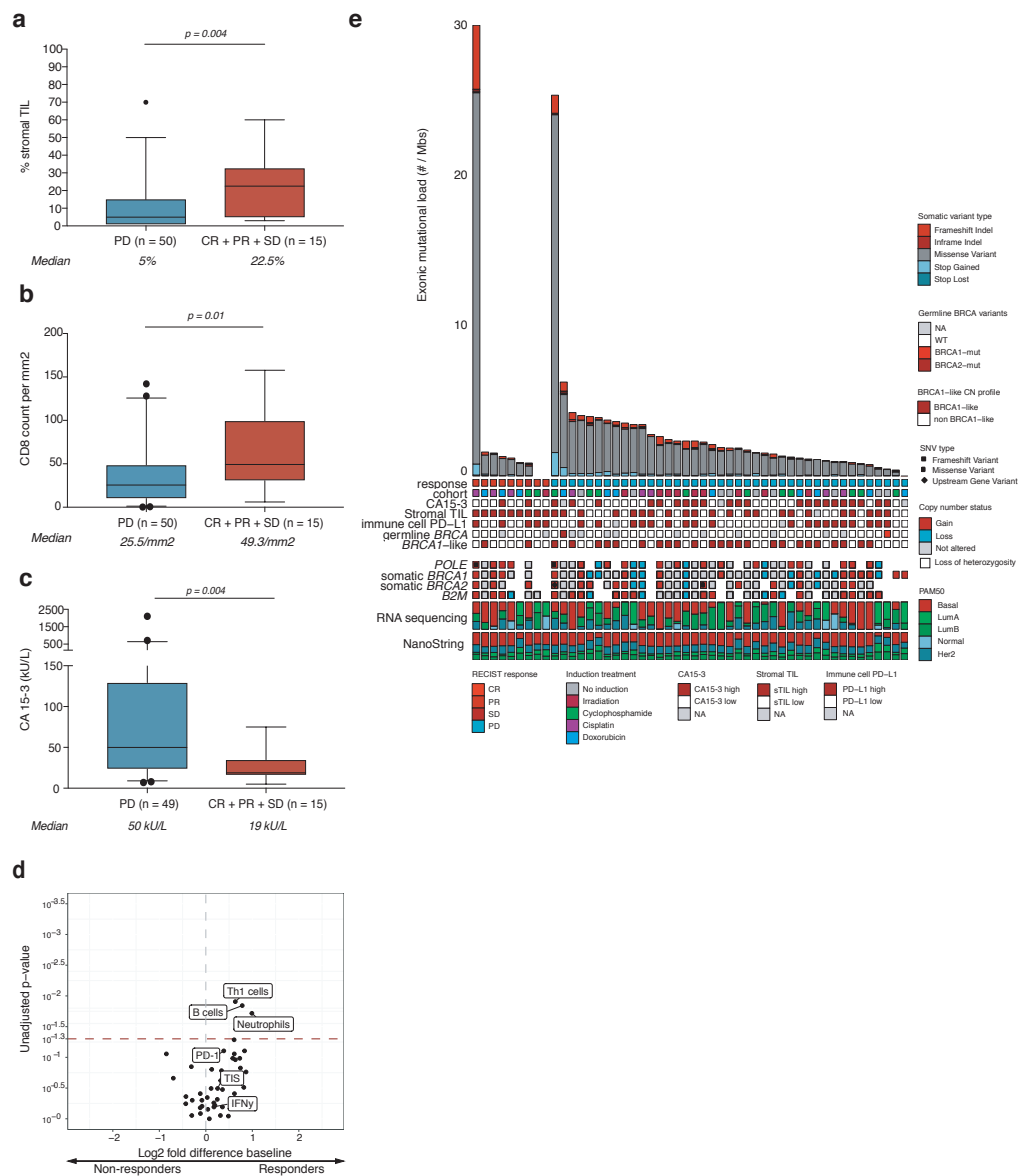


Figure 3.2: Intratumoral and systemic baseline parameters associated with response.

a Baseline sTILs determined according to guidelines of the TIL working group on a H&E staining of tumor biopsies. The median value is displayed for patients with or without clinical benefit; the median in the overall

cohort was 5%. Boxplots represent the median and 25th and 75th percentiles, and the vertical bars span the 5th to the 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* value).

b Baseline CD8 cell count per mm² in tumor biopsies. The median value is displayed for patients with or without clinical benefit; the median in the overall cohort was 30 cells per mm². Boxplots represent the median and 25th and 75th percentiles, and the vertical bars span the 5th to the 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* value).

c Baseline serum levels of CA 15-3. CA 15-3 was measured according to local guidelines. The median value is displayed for patients with or without clinical benefit; the median in the overall cohort was 35 kU l⁻¹ (which is 1× ULN). Boxplots represent the median and 25th and 75th percentiles, and the vertical bars span the 5th to the 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* value).

d Volcano plot of baseline gene expression signatures assessed with the NanoString IO 360 panel of 770 genes. Displayed is the log₂ fold difference of the median gene expression signature score between non-responders and responders (all patients with clinical benefit). Statistical significance is observed for signatures above the red dashed line (two-sided Wilcoxon signed-rank test; unadjusted *P* value of 0.05). Every dot represents one gene expression signature, as previously determined by Ayers et al.²² and Danaher et al.^{23,26}. The gray dashed line indicates no difference in gene expression. IFN- γ , interferon- γ ; T_H1, T helper 1; TIS, tumor inflammation signature.

e Mutational load, germline (according to routine clinical diagnostics) and somatic *BRCA* variants, *BRCA1*-like copy number (CN) profiles, copy number or mutation status of *POLE*, *BRCA1*, *BRCA2* and *B2M*, and PAM50 subtype assessed by RNA sequencing and NanoString are depicted. Data were available for 50 patients, samples were taken at baseline before study treatment. NA, not available; SNV, single-nucleotide variant; WT, wild type; LumA, Luminal A; LumB, Luminal B.

We explored the potential predictive value of clinical characteristics and baseline aspects of the TME and peripheral blood. Patients with a disease-free interval (DFI) of 1 year or shorter had lower response rates ($p = 0.02$; Figure 3.2a). The ORR for patients treated in the first line was 33%, while the ORR was 16% in patients treated in the second or later lines ($p = 0.15$; Figure S3.2a). We observed significantly higher levels of stromal tumor-infiltrating lymphocytes (sTILs) and higher levels of CD8 and PD-L1 on immune cells in responders than in non-responders (Figure 3.2a,b) and Figure S3.2a,b and 3.3). Furthermore, we observed significantly lower cancer antigen 15-3 (CA 15-3) and carcinoembryonic antigen (CEA) levels in responders (Figure 3.2c and Figure S3.2a,d). CA 15-3 showed a moderate correlation with the number of metastatic sites (Figure S3.2e). In a multivariate analysis, CA 15-3 remained associated with response after adjustment for sTILs and lines of treatment (odds ratio: 0.69; $p = 0.05$) but not after adjustment for number of metastatic sites (odds ratio: 0.72; $p = 0.08$). No significant correlation with response was observed for lactate dehydrogenase (LDH), C-reactive protein, neutrophils, lymphocytes, neutrophil-to-lymphocyte ratio, eosinophils or serum levels of 12 CD8 T cell and natural-killer-cell-related cytokines (Figure S3.2f–k and S3.4). In addition, we observed higher gene signature scores for T helper 1 cells, B cells and neutrophils in responders than in non-responders (Figure 3.2d), using the NanoString IO 360 Panel. Higher T cell receptor (TCR) clonality, more T cells and a larger TCR repertoire diversity (the number of unique intratumoral T cell clones) were observed in responders than in non-responders, both intratumoral and in the blood (Figure S3.2l–q), however these associations were not statistically significant. We found no association between mutational load or predicted neo-epitopes and response (Figure 3.2e

and Figure S3.2r,s). Two patients with exceptionally high mutational loads had somatic mutations in *POLE*. One of those cases was also identified as microsatellite instable and had a durable response²⁰. No mutations in *B2M* were observed at baseline. We found no associations between response and *BRCA1/2* mutations (Figure 3.2e), but we did observe *BRCA1*-like genomic copy number profiles to be negatively associated with response (Figure 3.2e and Figure S3.2a). Molecular subtypes according to PAM50²¹ were not associated with response.

Most clinical responses occurred during nivolumab, with two patients having the onset of response during the induction period (Figure S3.5a,b). Most responses were observed in the doxorubicin cohort (ORR: 35%; 95% CI: 14.2–61.7%), followed by cisplatin (ORR: 23%; 95% CI: 5–53.8%; Figure 3.1c,d). In the no induction treatment cohort, two patients experienced a PR (ORR: 17%; 95% CI: 2.1–48.4%); in the irradiation and cyclophosphamide cohorts, only one patient had a PR (ORR: 8%; 95% CI: 0.2–38.5%). When restricting analysis to non-first-line patients, the doxorubicin, cisplatin and no inductions cohorts still showed numerically higher ORRs than the irradiation and cyclophosphamide cohorts (Figure S3.5c). According to the Simon’s two-stage design²², discontinuation of a cohort was required if fewer than three out of ten patients had no progressive disease after 12 weeks (Figure S3.5e–g). According to iRECIST (Figure S3.5e,f), only the doxorubicin cohort was allowed to continue.

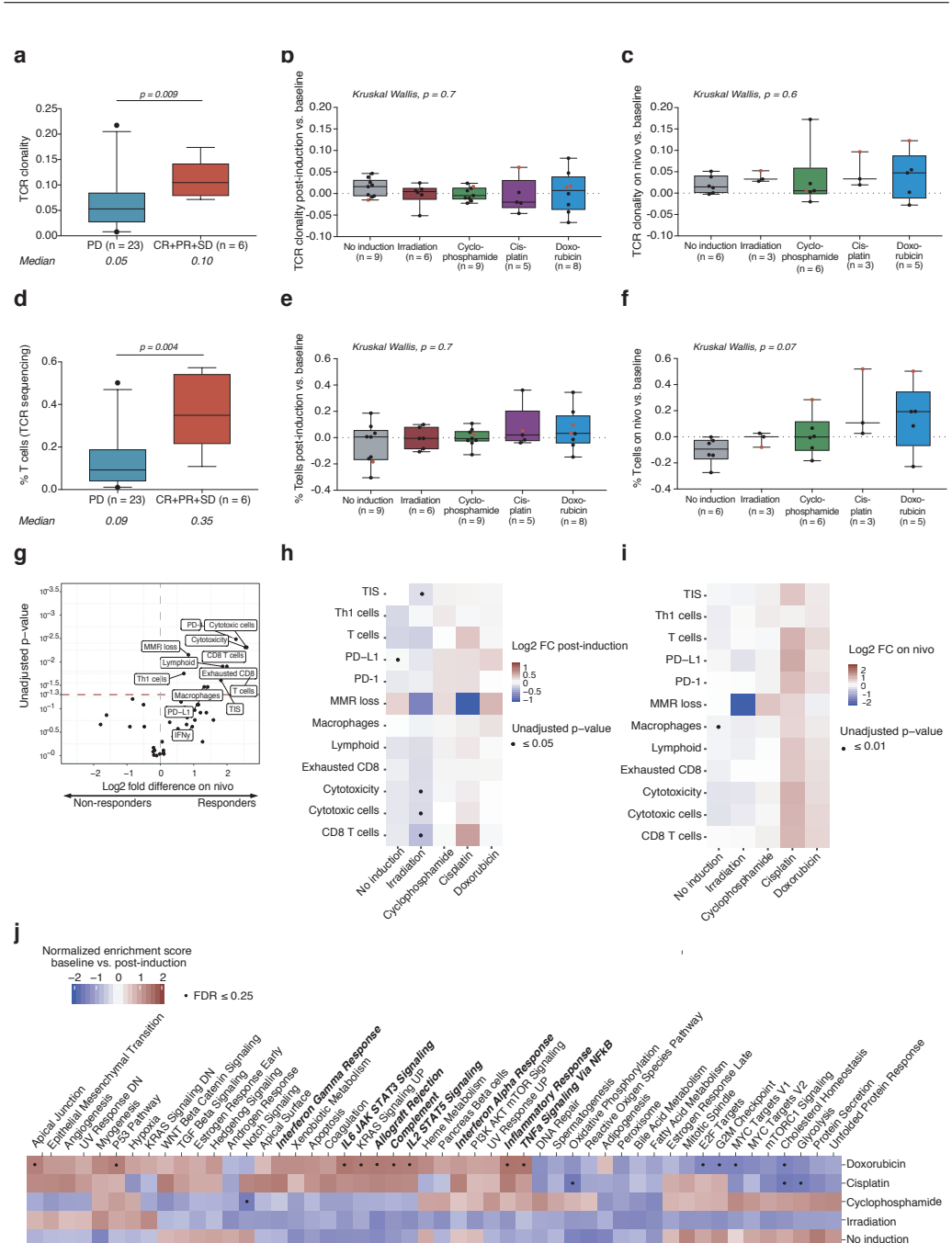


Figure 3.3: Characteristics of tumors during an active ongoing anticancer response on nivolumab and changes observed after induction treatments

a Intratumoral TCR clonality on nivolumab treatment. On nivolumab treatment, TCR sequencing data were available for 29 patients. Boxplots represent the median, 25th and 75th percentiles and the vertical bars span the 5th to the 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* value).

b Fold change (FC) in TCR clonality after induction treatment versus baseline (biopsy two versus biopsy one).

The clonality of TCRs indicates the specific expansion of a subset of T cell clones. The boxes in the boxplots represent the median and interquartile ranges, and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change.

c FC in TCR clonality on nivolumab (nivo) treatment versus baseline (biopsy three versus biopsy one). The boxes in the boxplots represent the median and interquartile ranges, and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change.

d Percentage of tumor-infiltrating T cells (TCR sequencing) on nivolumab treatment. The percentage of T cells of total nucleated cells, as assessed by TCR sequencing is depicted. Boxplots represent the median, 25th and 75th percentiles and the vertical bars span the 5th to the 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* value).

e FC in the percentage of tumor-infiltrating T cells (TCR sequencing) after induction treatment versus baseline (biopsy two versus biopsy one). The percentage of T cells over nucleated cells is depicted. The boxes in the boxplots represent the median and interquartile ranges and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change.

f FC in the percentage of tumor-infiltrating T cells (TCR sequencing) on nivolumab treatment versus baseline (biopsy three versus biopsy one). The percentage of T cells over nucleated cells is depicted. The boxes in the boxplots represent the median and interquartile ranges and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change.

g Volcano plot of previously established gene expression signatures^{22,23,26}, assessed with the NanoString IO 360 panel of 770 genes. The log₂ fold difference of the median gene expression per signature between non-responders and patients with clinical benefit in biopsies taken after three cycles of nivolumab (biopsy three) is displayed. Statistical significance is observed for signatures above the red dashed line (two-sided Wilcoxon signed-rank test; unadjusted *P* value of 0.05). Each dot represents one of the previously established gene expression signatures by NanoString^{22,23,26}. The gray dashed line indicates no difference in gene expression. MMR, DNA mismatch repair.

h Heatmap of post-induction FCs in gene expression signatures (NanoString; significantly upregulated during an active ongoing response on nivolumab, determined in **g**) in post-induction samples (biopsy two) compared to baseline (biopsy one). Depicted is the log₂ FC in median gene expression of paired biopsies. Statistical significance (two-sided Wilcoxon signed-rank test) is highlighted with a black dot.

i Heatmap of on-nivolumab FCs in gene expression signatures (NanoString; significantly upregulated during an active ongoing response on nivolumab, determined in **g**) in samples taken on nivolumab (biopsy three) compared to baseline (biopsy one). Depicted is the log₂ FC in median gene expression of paired biopsies. Statistical significance (two-sided Wilcoxon signed-rank test) is highlighted with a black dot.

j Gene set enrichment analysis of 50 hallmark gene sets²⁴, performed on whole-transcriptome RNA sequencing of pre-induction and post-induction samples (biopsy 2 compared to biopsy 1). Cells are colored according to normalized enrichment scores, and Benjamini–Hochberg (FDR) corrected *P* values equaling or below 0.25 are highlighted with black dots. Immune-related gene sets are highlighted in bold font. DN, downregulated; IL-6, interleukin-6; JAK, Janus kinase; mTOR, mechanistic target of rapamycin; mTORC1, mTOR complex 1; NF- κ B, nuclear factor- κ B; PI3K, phosphatidylinositol-3-OH kinase; STAT3, signal transducer and activator of transcription 3; TGF- β , transforming growth factor- β ; TNF- α , tumor necrosis factor- α ; UP, upregulated.

The main objective of the TONIC trial was to explore whether induction treatment can induce a more inflamed TME. To determine the desired state of inflammation, we first studied the ongoing anti-cancer immune response in biopsy three of responders compared to biopsy three from non-responders. On nivolumab, we observed higher TCR clonality ($p = 0.009$) and increased T cell infiltra-

tion ($p = 0.004$; Figure 3.3a,d). Although T cell repertoire clonality appeared more strongly increased in the cisplatin and doxorubicin cohorts after nivolumab treatment (biopsy three versus biopsy one) than in the control cohort (Figure 3.3c), such inter-cohort differences were not observed directly after induction (biopsy two versus biopsy one; Figure 3.3b). In addition, we observed a trend in increased T cell infiltration after induction with cisplatin and doxorubicin (biopsy two versus biopsy one; Figure 3.3e), which became more pronounced after nivolumab treatment (biopsy three versus biopsy one; Figure 3.3f). Finally, increases in the number of unique intratumoral T cell clones (TCR diversity) were significantly higher on nivolumab in the doxorubicin cohort than in the control cohort (Figure S3.6). We observed higher TIL (H&E) and CD8 counts (immunohistochemistry) in on-nivolumab biopsies of responders than in non-responders (Figure S3.7a,b). Comparing post-induction and baseline, we observed a trend towards increased TIL and CD8 counts in all cohorts except for the irradiation cohort (Figure S3.7c,d) and non-significant increases in TIL and CD8 counts after nivolumab treatment in the doxorubicin cohort. We observed no changes in stromal CD4 or FOXP3 expression after induction. A non-significant increase in CD4 expression in the doxorubicin cohort was observed (Figure S3.7e,f).

Next, we evaluated treatment-induced changes in the expression of immune-related genes (NanoString IO 360 Panel)^{22,23}. On nivolumab (biopsy three), several gene signatures associated with inflammation were significantly higher for responders than for non-responders (Figure 3.3g). Following cisplatin and doxorubicin treatments, most of these inflammation-related signatures (Figure 3.3g) showed a trend towards upregulation, but after irradiation or a 2-week waiting period these signatures tended to get downregulated (biopsy two versus biopsy one; Figure 3.3h). Upregulation of inflammation-related signatures in the cisplatin and doxorubicin cohorts was even more pronounced after nivolumab treatment (biopsy three versus biopsy one; Figure 3.3i). Using a Bayesian model, we estimated the effect sizes of the four induction treatments on immune-related gene signatures (Figure 3.3g). We observed that the effect sizes of cisplatin and doxorubicin equaled or exceeded changes in the no induction cohort with 98.0% and 85.2% probability, respectively (Figure S3.8b). After correction for baseline gene expression, clinical response to nivolumab, lines of palliative treatment and lymph node only metastasis, probabilities of 92.1% and 80.7% (Figure S3.8g,h), respectively, were obtained. Subsequently, a gene set enrichment analysis (GSEA) on 50 hall-mark gene sets²⁴ on RNA sequencing data demonstrated an enrichment of eight immune-related gene sets among upregulated genes (biopsy two versus biopsy one) after doxorubicin treatment (six out of eight gene sets passed multiple testing correction) and after cisplatin treatment (zero out of eight passed multiple testing correction). After irradiation and cyclophosphamide treatments, the majority of these gene sets showed a non-significant enrichment among down-regulated genes. By contrast, only 7 out of 42 non-immune-related hallmark gene sets were enriched among upregulated genes after doxorubicin. In addition, we tested previously established gene signatures related to myeloid cells^{23,25,26} and CD4 T cells²⁷. Three (out of four) myeloid-related signatures showed upregulation after induction and/or on nivolumab treatment (Figure S3.9a,b). Furthermore, we evaluated two MDSC-related signatures²⁵ and two CD4 T cell signatures²⁷ in a separate GSEA and observed all to be enriched

among upregulated genes after doxorubicin and cisplatin (false discovery rate ≤ 0.25 ; Figure S3.9c).

To our knowledge, TONIC is the first trial to evaluate the concept of TME modulation using chemotherapy or irradiation prior to PD-1/PD-L1 blockade. Our data provide substantial evidence that induction with cisplatin or doxorubicin can prime tumors for response to anti-PD-1, based on high response rates to anti-PD-1 and upregulation of immune-related gene sets. Finally, we observed trends towards increased T cell infiltration and TCR diversity in the doxorubicin cohort. On the basis of the Simon's two-stage design, the doxorubicin cohort is currently expanded in stage II of the trial (Figure S3.5h). We note that this trial was not designed nor powered for direct comparison of response rates between arms and, as such, the data cannot be used as conclusive evidence for the inferiority of other induction treatments.

The majority of clinical trials that evaluate immune checkpoint blockade (ICB) in combination with chemotherapy simply combine PD-1/PD-L1 blockade with standard chemotherapy^{5,28}, which was shown to lead to increased survival for patients with PD-L1-positive TNBC⁵. By contrast, the sequential administration of chemotherapy or irradiation in the TONIC trial allowed us to test whether conventional treatments can turn 'cold' into 'hot' tumors. To the best of our knowledge, strong preclinical or clinical data that assess whether the sequential use of chemotherapy or irradiation is better than concomitant use are still lacking. Arguments in favor of the latter are the relatively long time to response to PD-1/PD-L1 blockade during which chemotherapy can provide early tumor control and the potential synergy between PD-1/PD-L1 blockade and chemotherapy. Conversely, the short-term use of chemotherapy reduces toxicity substantially while potentially still effectuating the immunomodulation associated with cytostatic agents. Interestingly, the first results of the GeparNuevo trial, evaluating anti-PD-L1 added to chemotherapy in primary TNBC, suggested that induction with anti-PD-L1 increased responses in primary TNBC²⁹. Of note, our analyses of sequential on-treatment biopsies revealed that the immunomodulatory effects induced by three cycles of anti-PD-1 were substantially larger than the changes observed after 2 weeks of induction, arguing for an earlier start of ICB. Recent work has demonstrated that high response rates are observed upon ICB in the neoadjuvant setting in melanoma and non-small-cell lung cancer^{30,31}, supporting the notion that primary tumors may be more sensitive to immune control than metastases. Given this, it would be interesting to apply the design of TONIC to the neoadjuvant setting.

We found that nivolumab in patients with metastatic TNBC resulted in an ORR of 20%. This ORR is higher than in other studies in this patient population that show an ORR of only 5–10%^{3,4}. This may be due to the priming strategy that was used in our trial, but patient selection may also have contributed, for example, as, in contrast to some previous studies^{3,4}, we excluded patients with high serum levels of LDH. Importantly, we confirm that patients with a short DFI (<1 year) had a low likelihood (<5%) of response to ICB even when the LDH level is <2× upper limit normal (ULN), as previously reported¹⁸. In contrast to data for melanoma and non-small-cell lung cancer^{32,33}, the tumor mutational burden did not correlate with response in our series, in line with previous work^{34,35}.

Although this lack of correlation may simply be explained by small sample sizes, it is interesting to consider that the tumor mutational burden might not be a determinant of response in relatively ‘cold’ tumors, such as breast cancer, in which copy number aberrations are more prevalent. This is supported by the significant association that we observed between *BRCA1*-like copy number profiles and non-response to anti-PD-1 and in line with previous studies in melanoma suggesting that copy number aberration burden is negatively associated with anti-PD-1 response^{36,37}.

We observed a significant correlation between PD-L1 on immune cells and nivolumab benefit (Figure 3.2a,b and 3.3e,f), in line with several studies in TNBC³⁻⁵. Of note, the prevalence of PD-L1-positive tumors is somewhat higher in our study than in other studies^{2,18}, with 86% of patients expressing PD-L1 on immune cells (assessed using the 22C3 clone). Using the SP142 clone and by scoring of immune cells, Schmid et al.⁵ reported 41% PD-L1-positive tumors in the first-line setting, whereas Emens et al.³ reported 81% PD-L1 positivity in heavily pretreated patients. Studies in non-small-cell lung cancer³⁸ and bladder cancer³⁹ have shown that the 22C3 clone yields higher proportions of PD-L1 positivity than the SP142 assay.

Although the TONIC trial design allowed relatively quick prioritization of treatments, there are several limitations to address. First, the TONIC trial was designed as a non-comparative trial with relatively small numbers. Although we only included patients with TNBC, this population is still quite heterogeneous in terms of previous treatments and metastatic patterns. Second, no stratification was applied in the first stage of the trial. Consequently, the cohorts were not perfectly balanced for several characteristics, such as the location of metastases and the number of previous palliative treatments. Third, we required a short period of preferential recruitment to the doxorubicin arm ($n = 17$) to obtain at least ten good-quality paired biopsies. As such, we cannot exclude that low-dose doxorubicin might also have a direct anti-tumor effect. Finally, according to the very stringent decision rules (requiring a success rate of at least 30%) that we set before the start of the trial (2014) without knowing that the ORR to PD-1/PD-L1 blockade would be only 5–10%^{1,3,4}, doxorubicin was picked as a winner only when the iRECIST criteria (ORR 35%) were applied, but not according to RECIST1.1⁴⁰ (Figure S3.5e–g).

In summary, induction treatment with short-term chemotherapy or irradiation followed by nivolumab is feasible and leads to clinical benefit in a substantial subset of patients, with higher than expected response rates and durable responses. Priming with doxorubicin or cisplatin seems to induce a more favorable TME and was associated with a higher likelihood of response to nivolumab in this study. Next to the ongoing validation in stage II of this TONIC trial, which incorporates a nivolumab monotherapy cohort and a doxorubicin followed by nivolumab cohort (Figure S3.5h), independent validation of our findings is required. In addition, the design of this study may serve as a template for other signal-finding combination immunotherapy studies in breast cancer and beyond.

Bibliography

1. Adams, S., *et al.* Pembrolizumab Monotherapy for Previously Treated Metastatic Triple-Negative Breast Cancer: Cohort A of the Phase 2 KEYNOTE-086 Study. *Annals of Oncology*, mdy517 (2018).
2. Nanda, R., *et al.* Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. *Journal of Clinical Oncology* **34**, 2460-2467 (2016).
3. Emens, L.A., *et al.* Long-term Clinical Outcomes and Biomarker Analyses of Atezolizumab Therapy for Patients With Metastatic Triple-Negative Breast Cancer: A Phase 1 Study. *JAMA oncology* **5**, 74-82 (2018).
4. Dirix, L.Y., *et al.* Avelumab, an anti-PD-L1 antibody, in patients with locally advanced or metastatic breast cancer: a phase 1b JAVELIN Solid Tumor study. *Breast cancer research and treatment* **167**, 671-686 (2018).
5. Schmid, P., *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *The New England Journal of Medicine* **380**, 987-988 (2018).
6. Demaria, S., *et al.* Ionizing radiation inhibition of distant untreated tumors (abscopal effect) is immune mediated. *International Journal of Radiation Oncology*Biology*Physics* **58**, 862-870 (2004).
7. Vanpouille-Box, C., *et al.* DNA exonuclease Trex1 regulates radiotherapy-induced tumour immunogenicity. *Nature communications* **8**, 15618 (2017).
8. Scurr, M., *et al.* Low-Dose Cyclophosphamide Induces Antitumor T-Cell Responses, which Associate with Survival in Metastatic Colorectal Cancer. *Clinical cancer research* **23**, 6771-6780 (2017).
9. de Biasi, A.R., Villena-Vargas, J. & Adusumilli, P.S. Cisplatin-induced antitumor immunomodulation: a review of preclinical and clinical evidence. *Clinical cancer research* **20**, 5384-5391 (2014).
10. Wan, S., *et al.* Chemotherapeutics and radiation stimulate MHC class I expression through elevated interferon-beta signaling in breast cancer cells. *PloS one* **7**, e32542 (2012).
11. Alizadeh, D., *et al.* Doxorubicin eliminates myeloid-derived suppressor cells and enhances the efficacy of adoptive T-cell transfer in breast cancer. *Cancer Research* **74**, 104-118 (2014).
12. Sistigu, A., *et al.* Cancer cell-autonomous contribution of type I interferon signaling to the efficacy of chemotherapy. *Nature medicine* **20**, 1301-1309 (2014).
13. Casares, N., *et al.* Caspase-dependent immunogenicity of doxorubicin-induced tumor cell death. *The Journal of experimental medicine* **202**, 1691-1701 (2005).

14. Seymour, L., *et al.* iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *The Lancet Oncology* **18**, e143-e152 (2017).
15. O'Brien, K.M., *et al.* Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clinical cancer research* **16**, 6100-6110 (2010).
16. den Brok, W.D., *et al.* Survival with metastatic breast cancer based on initial presentation, de novo versus relapsed. *Breast cancer research and treatment* **161**, 549-556 (2017).
17. Kassam, F., *et al.* Survival outcomes for patients with metastatic triple-negative breast cancer: implications for clinical practice and trial design. *Clinical breast cancer* **9**, 29-33 (2009).
18. Adams, S., *et al.* Title: Pembrolizumab Monotherapy for Previously Untreated, PD-L1-Positive, Metastatic Triple-Negative Breast Cancer: Cohort B of the Phase 2 KEYNOTE-086 Study. *Annals of Oncology* **mdy518** (2018).
19. Simon, R. Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials* **10**, 1-10 (1989).
20. Kok, M., *et al.* Profound Immunotherapy Response in Mismatch Repair-Deficient Breast Cancer. *JCO Precision Oncology* **1**, 1-3 (2017).
21. Sørlie, T., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10869-10874 (2001).
22. Ayers, M., *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *Journal of Clinical Investigation* **127**, 2930-2940 (2017).
23. Danaher, P., *et al.* Gene expression markers of Tumor Infiltrating Leukocytes. *Journal for immunotherapy of cancer* **5**, 18 (2017).
24. Liberzon, A., *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417-425 (2015).
25. Bezzi, M., *et al.* Diverse genetic-driven immune landscapes dictate tumor progression through distinct mechanisms. *Nature medicine* **24**, 165-175 (2018).
26. Danaher, P., Warren, S. & Cesano, A. Development of gene expression signatures characterizing the tumor-immune interaction. *Journal of Clinical Oncology* **36**, 205 (2018).
27. Gu-Trantien, C., *et al.* CD4+ follicular helper T cell infiltration predicts breast cancer survival. *Journal of Clinical Investigation* **123**, 2873-2892 (2013).
28. Adams, S., Diamond, J.R., Hamilton, E. & *et al.* Atezolizumab plus nab-paclitaxel in the treatment of metastatic triple-negative breast cancer with 2-year survival follow-up: A phase 1b clinical trial. *JAMA oncology* **5**, 334-342 (2019).

-
29. Loibl, S., *et al.* Randomized phase II neoadjuvant study (GeparNuevo) to investigate the addition of durvalumab to a taxane-anthracycline containing chemotherapy in triple negative breast cancer (TNBC). *Journal of Clinical Oncology* **36**, 104 (2018).
 30. Blank, C.U., *et al.* Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. *Nature medicine* **24**, 1655-1661 (2018).
 31. Forde, P.M., *et al.* Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *The New England Journal of Medicine* **378**, 1976-1986 (2018).
 32. Rizvi, N.A., *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (New York, N.Y.)* **348**, 124-128 (2015).
 33. Van Allen, E.M., *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science (New York, N.Y.)* **350**, 207-211 (2015).
 34. Samstein, R.M., *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature genetics* **51**, 202-206 (2019).
 35. Molinero, L., *et al.* Abstract SABCS 2017 P2-09-13: Molecular characterization of tumors from metastatic TNBC patients treated with atezolizumab (atezo). **78**, P2-09-13 (2018).
 36. Davoli, T., Uno, H., Wooten, E.C. & Elledge, S.J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (New York, N.Y.)* **355** eaaf8399 (2017).
 37. Roh, W., *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Science translational medicine* **9** eaah3560 (2017).
 38. Rimm, D.L., *et al.* A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. *JAMA oncology* **3**, 1051-1058 (2017).
 39. Zavalishina, L., *et al.* RUSSCO-RSP comparative study of immunohistochemistry diagnostic assays for PD-L1 expression in urothelial bladder cancer. *Virchows Archiv* **473**, 719-724 (2018).
 40. Eisenhauer, E.A., *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European journal of cancer (Oxford, England : 1990)* **45**, 228-247 (2009).
 41. Baselga, J., *et al.* Randomized phase II study of the anti-epidermal growth factor receptor monoclonal antibody cetuximab with cisplatin versus cisplatin alone in patients with metastatic triple-negative breast cancer. *Journal of Clinical Oncology* **31**, 2586-2592 (2013).
 42. Singh, J., *et al.* Phase 2 trial of everolimus and carboplatin combination in patients with triple negative metastatic breast cancer. *Breast cancer research* **16**, R32 (2014).
 43. O'Shaughnessy, J., *et al.* Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *The New England Journal of Medicine* **364**, 205-214 (2011).

44. Carey, L.A., *et al.* TBCRC 001: randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. *Journal of Clinical Oncology* **30**, 2615-2623 (2012).
45. R Core Team R: a language and environment for statistical computing v3.3.2 (2018).
46. Cibulskis, K., *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* **31**, 213–219 (2013).
47. Favero, F., *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology* **26**, 64-70 (2015).
48. Love, M.I., Hogenesch, J.B. & Irizarry, R.A. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature Biotechnology* **34**, 1287-1291 (2016).
49. Nielsen, M., Lundegaard, C., Lund, O. & Kesmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33-41 (2005).
50. Kesmir, C., Nussbaum, A.K., Schild, H., Detours, V. & Brunak, S. Prediction of proteasome cleavage motifs by neural networks. *Protein engineering* **15**, 287-296 (2002).
51. Jurtz, V., *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *Journal of immunology* **199**, 3360-3368 (2017).
52. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
53. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research* **40**, 4288-4297 (2012).
54. Efron, B. & Thisted, R. Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63**, 435-447 (1976).
55. Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545-15550 (2005).
56. Gendoo, D.M., *et al.* Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097-1099 (2016).
57. Carpenter, B., *et al.* Stan: A Probabilistic Programming Language. *J. Stat. Softw* **76**, 1-32 (2017).
58. Joosse, S.A., *et al.* Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. *Breast cancer research and treatment* **116**, 479-489 (2009).
59. Schouten, P.C., *et al.* Robust BRCA1-like classification of copy number profiles of samples repeated across different datasets and platforms. *Molecular oncology* **9**, 1274-1286 (2015).

-
60. Vollebergh, M.A., *et al.* An aCGH classifier derived from BRCA1-mutated breast cancer and benefit of high-dose platinum-based chemotherapy in HER2-negative breast cancer patients. *Annals of Oncology* **22**, 1561-1570 (2011).

Acknowledgements

We thank the patients and their families for participating in the study. We thank J. Foekema, M. Holtkamp, M. Delfos, J. van Zyl-de Jong and K. Kersten for their support in the care for patients. We thank S. Vanhoutvin for legal support. We thank the Core Facility of Molecular Pathology & Biobanking for their support in processing of samples. In addition, we acknowledge the Genomics Core Facility for their support regarding sequencing. We acknowledge J. Lips from Adaptive Biotechnologies for his support. We thank the scientific administration department, in particular L. Ruiter, for data management/monitoring. The Clinical Chemistry Department is thanked for their support in blood withdrawals. We thank H. Garner, M. van der Heijden and J. Stouthard for critical reading of the manuscript. We acknowledge D. Cullen, A. Evans, D. Zardavas and D. Feltquate of Bristol-Myers Squibb (BMS) for scientific input. We thank BMS/II-ON and the Dutch Cancer Society (NKI2015–7710, 10653 ALPE) for funding the study and a fellowship to M.K. (NKI2015–7542). Pink Ribbon (NKI2016–8214), the Breast Cancer Research Foundation (BCRF-17–188) and BMS/II-ON are thanked for the funding of the translational research. R.S. is supported by a grant from the Breast Cancer Research Foundation (BCRF-17–194).

Author contributions

L.V. coordinated trial procedures, analyzed and interpreted clinical and translational data and wrote the manuscript with M.S., T.N.S., C.U.B. and M.K. M.S. performed and interpreted the bioinformatic analyses. H.M.H., K.K.v.d.V. and R.S. performed the histological scoring. K.S. performed the statistical analysis on the clinical data. M.d.M. was responsible for DNA and RNA isolations. I.N. provided input during work discussions. R.J.C.K. processed the raw DNA and RNA sequencing data. S. Warren and S. Ong were responsible for the NanoString nCounter assay experiments and analyses. T.G.W. and N.S.R. were responsible for initial screening and the patients treated with irradiation. F.L. revised the CT scans. P.C.S. adapted the *BRCA1*-like classifier and applied it to our data set. N.A.M.B. and L.V. performed and analyzed the cytokine assays. S.L.C.K. performed the prediction of neo-epitopes. D.P. was responsible for the double staining of CD4 and FOXP3. C.A.H.L. performed the majority of biopsies and assessment of the CT scans. E.v.W. and H.v.T. were involved in the statistical design. I.A.M.M. was the clinical projects manager involved in the trial. I.K. and S. Onderwater. were responsible for patient care. M.C., S. Wilgenhof, G.S.S., S.C.L. and M.K. included patients in the trial and were responsible for patient care. J.B.A.G.H. advised on the trial design. K.E.d.V. gave critical input and supervised the cytokine assays. L.F.A.W. supervised the bioinformatics analyses. G.S.S., S.C.L., C.U.B., T.N.S. and M.K. designed the trial. C.U.B., T.N.S. and M.K. made

the experimental plan of investigation. All authors edited and approved the manuscript.

Competing interests

L.V., M.S., H.M.H., K.S., K.K.v.d.V., M.d.M., I.N., R.J.C.K., T.G.W., N.S.R., F.L., N.A.M.B., S.L.C.K., D.P., C.A.H.L., E.v.W., H.v.T., I.A.M.M., I.K., S. Onderwater and S. Wilgenhof declare no competing interests. S.Warren reports employment and stockholdership of NanoString Technologies, an advisory role for Roche and being a former employee of Oncofactor Corp., outside the submitted work. S. Ong reports employment and stockholdership of NanoString Technologies. P.C.S. has a close relative employed by AstraZeneca. M.C. reports funding to the institute from BMS and Roche/Genentech, outside the submitted work. J.B.A.G.H. reports financial compensation to the NKI for advisory roles from Amgen, AZ, BMS, Bayer, MSD, Celsius Therapeutics, Gadeta, Immunocore, Seattle Genetics, Merck Serono, Sanofi, Roche, Neon therapeutics, Pfizer and Ipsen and NKI, and received grants from BMS, MSD, Novartis and Neon therapeutics, outside the submitted work. R.S. reports research funding from Merck, Roche and Puma, as well as travel funds from AstraZeneca, Roche, Merck and an advisory role for BMS, outside the scope of this work. K.E.d.V. reports research funding from Roche, outside the scope of this work. G.S.S. reports funding to the institute from AstraZeneca, Merck Sharp & Dohme, Novartis and Roche, outside the submitted work. L.F.A.W. reports receiving a commercial research grant from Genmab. S.C.L. reports funding to the institute from Agenzia, Amgen, AstraZeneca, BMS, Eurocept, Roche/Genentech, Tesaro and an advisory role for AstraZeneca, Bayer and IBM, outside the submitted work. T.N.S. is a consultant for Adaptive Biotechnologies, AIMM Therapeutics, Allogene Therapeutics, Amgen, Merus, Neon Therapeutics, Scenic Biotech; received grant or research support from Merck, BMS and Merck KGaA; is a stockholder in AIMM Therapeutics, Allogene Therapeutics, Neon Therapeutics and Neogene Therapeutics, all outside the submitted work. C.U.B. reports personal fees for advisory roles for MSD, BMS, Roche, GSK, Novartis, Pfizer, Lilly, Pierre Fabre and GenMab and grants from BMS, Novartis and NanoString, outside the submitted work. M.K. reports funding and a speaker's fee to the institute from BMS and Roche and an unpaid advisory role for BMS, outside the submitted work.

Methods

Study design

The TONIC trial (full title: adaptive phase 2 randomized non- comparative trial of nivolumab after induction treatment in triple-negative breast cancer patients; NCT02499367) is a single center, non-blinded, randomized, non-comparative phase II study designed to evaluate the feasibility and efficacy of nivolumab after a 2-week induction treatment with chemotherapy or irradiation in patients with metastatic TNBC. The first stage of the trial consisted of five cohorts (four with induction treatment

before nivolumab, one with a 2-week waiting period), all with a Simon's two-stage design¹⁹. For the second stage, the number of arms is reduced based on the results obtained in the first stage, according to the 'pick-the-winner' principle, considering clinical as well as translational end points. The trial was conducted in accordance with the protocol, Good Clinical Practice standards and the Declaration of Helsinki. The full protocol, including two amendments, and the informed consent form were approved by the institution's medical-ethical committee. All patients provided written informed consent before enrollment. This investigator-initiated trial was designed by the Netherlands Cancer Institute (NKI). Funding was provided by Bristol-Myers-Squibb (BMS) through the International Immuno-Oncology Network (II-ON) and by the Dutch Cancer Society (NKI2015-7710) with the NKI being the sponsor. Translational research was funded by Pink Ribbon (NKI2016-8214), the Breast Cancer Research Foundation (BCRF-17-188) and BMS/II-ON. The study protocol was written during the ECCO-AACR-ESMO-EORTC course 'Methods in Clinical Cancer Research' Flims, 2014.

Patients

Key inclusion criteria included: 18 years of age or older; metastatic or incurable locally advanced TNBC with confirmation of estrogen receptor and HER2 negativity (ER < 10% and HER2 0, 1 or 2 in the absence of amplification as determined by in situ hybridization) on a biopsy of a metastatic lesion or recurrence in the breast; a WHO (World Health Organization) performance status of 0 or 1; measurable or evaluable disease according to RECIST1.1⁴⁰; and a maximum of three previous lines of palliative systemic treatment. Key exclusion criteria included: a LDH level above 500 U l⁻¹ (>2× ULN); symptomatic brain metastasis (treated and stable brain metastasis were allowed); previous therapy with ICB; and active autoimmune disease or chronic infections. Patients were not selected based on PD-L1 expression and had to have an accessible lesion for sequential biopsies and a different lesion accessible for irradiation. Full eligibility criteria are listed in the Supplementary Note. At the start of the trial, PD-L1 was assessed using immunohistochemistry and was used for stratification of the first 17 patients. For logistical reasons and an unacceptable waiting time for patients due to this PD-L1 analysis, this stratification procedure was stopped.

Procedures

Before the start of the induction treatment (biopsy one), before the start of nivolumab (biopsy two) and after 6 weeks of nivolumab (biopsy three), a biopsy was taken from a metastatic lesion, preferably the same lesion throughout the study. In the case of irradiation as induction treatment, a biopsy was taken from a non-irradiated lesion. When a good-quality baseline biopsy (at least 100 invasive tumor cells) of a metastatic lesion or recurrence in the breast was obtained, subjects were randomly allocated to 1 of 4 induction treatments. Induction treatments consisted of irradiation of 1 metastatic lesion (3 fractions of 8 Gy within 10 weekdays after randomization), cyclophosphamide (50 mg orally

daily for 2 weeks), cisplatin (40 mg per m² intravenously weekly for 2 weeks) or doxorubicin (15 mg intravenously weekly for 2 weeks). A fifth control cohort was subjected to a 2-week waiting period. The irradiation was delivered to an accessible lesion, which was defined as a metastatic, preferably visceral, otherwise lymph node or bone, lesion at a distant location from the biopsy site. The radiation technique depended on the metastasis site (Supplementary Table S6). In general, the lesion was expanded with a 5-mm margin to acquire a planning target volume. Tumor coverage was assessed by the volume of the planning target volume receiving 95% of the prescribed dose. All patients underwent a second biopsy, after which nivolumab (3 mg per kg intravenously every 2 weeks) was given until disease progression according to iRECIST¹⁴ or until unacceptable toxicity. Accrual to a cohort was continued until ten patients were included who received at least one cycle of nivolumab, and for whom we were able to obtain a good-quality biopsy at baseline and after induction treatment. Twelve patients were allocated to the control or no induction cohort, 12 to the irradiation cohort, 13 to the cyclophosphamide cohort, 13 to

the cisplatin cohort and 17 to the doxorubicin cohort. Clinically stable patients with radiographic evidence of progressive disease according to RECIST1.1 were permitted to continue nivolumab treatment until radiographic confirmation of progressive disease on a second CT scan. When patients had an ongoing response after 12 months of treatment, nivolumab was allowed to be discontinued and reintroduced when progressive disease occurred. Dose modification for nivolumab was not permitted, but dose interruptions were allowed in case of (or suspicion of) toxicity. Safety was assessed every 2 weeks and included monitoring of AEs by clinical laboratory assessments and physical examinations. AEs were classified and graded per National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE), v4.03. Serious AEs were collected up to 30 d after the last nivolumab administration. Imaging was performed after the 2-week induction treatment period and thereafter every 6 weeks until 6 months, after which imaging was performed every 8 weeks. Best overall response, duration of response and the date of progression were assessed according to RECIST1.1 and iRECIST, investigator assessed. An independent radiologist with extensive experience with response assessment in patients treated with ICB reviewed the scans of the responding cases.

End points

The primary end point of the study was PFS, assessed from randomization (PFS1) to tumor progression or death from any cause as defined by RECIST1.1. Secondary end points of the study were ORR, defined as the percentage of patients with a best overall response of CR or PR according to RECIST1.1 and iRECIST; clinical benefit rate, defined as the percentage of patients with a best overall response of CR, PR and stable disease for 24 weeks, according to RECIST1.1 and iRECIST; PFS1 as defined by iRECIST; PFS, assessed from nivolumab treatment initiation (PFS2) to tumor progression or death from any cause as defined by RECIST1.1 and iRECIST; overall survival, defined as the time from nivolumab initiation to death from any cause; and the percentage of patients with toxicity according to NCI-CTCAE v4.03 and immune-related toxicity. Translational objectives included: the

effects of the induction treatments on the anticancer immune response evaluated using immune-related gene expression signatures; T cell influx determined using H&E and immunohistochemistry and TCR sequencing; and the exploration of putative predictive biomarkers.

Statistical analysis

For patients with metastatic TNBC, no first-line ‘standard’ therapies have been defined. Frequently used anticancer agents are capecitabine or taxanes. The median PFS with these therapies typically lies between 4–6 months. No ‘standard’ second-line therapy exists for patients with TNBC, but carboplatin (\pm gemcitabine), vinorelbine, capecitabine and taxanes are often used. On the basis of four phase 2 trials in TNBC allowing one or two previous lines of chemotherapy, a median PFS between 2 and 4 months was anticipated^{41–44}. Thus, the investigators considered a proportion of >30% of the patients having a PFS of at least 12 weeks as potentially interesting. The null hypothesis that the true PFS rate as a binary end point at 12 weeks is 30% was tested against an alternative of 50%. A Simon two-stage minimax design with a one-sided alpha of 15% and 85% power was also optimal with respect to the expected sample size. A sample size of ten evaluable patients in the first stage required early discontinuation of a particular treatment cohort if less than three out of ten patients were free of progression and alive at 12 weeks. Because the number of patients in each cohort is larger than ten (due to the collection of ten paired biopsies), the decision about discontinuation of a cohort was based on the first ten patients. A patient was considered evaluable when at least one cycle of nivolumab was administered and both the baseline biopsy (biopsy one) and the post-induction biopsy (biopsy two) were available for immunohistochemistry. PFS and OS were assessed in all patients who received at least one dose of nivolumab (per protocol population). The safety population consisted of all patients who started their allocated treatment. PFS, OS, duration of response and median follow-up were calculated from the date of randomization and estimated using the Kaplan–Meier method. The duration of response was calculated from the first date of response to the date of progression. Median time to response was calculated as the time between randomization and the first measured objective response in responding cases. The DFI was defined as the time between the diagnosis of the primary tumor or locoregional recurrence and the date of diagnosis of metastatic disease. Patients with de novo metastatic disease at diagnosis were excluded from the exploratory analysis testing the association between DFI and ORR. A binary logistic regression analysis was performed to assess the effect of CA 15-3 (per 10 units) on response after correction for possible confounding factors (one model corrected for the number of metastatic sites and another model corrected for TIL and previous lines of treatment). As the number of metastatic sites and CA 15-3 were correlated (Spearman’s ρ : 0.46; $P = 0.0001$), we tested for multicollinearity and found a variance inflation factor of 1.02, indicating no multicollinearity. The number of metastatic sites (1–2 versus 3 or more sites) and the number of previous lines of treatment (0 versus 1–3 lines) were included as categorical variables with the lowest category as a reference. Two-sided non-parametric tests were used for all analyses of the translational data: that is, the Mann–Whitney U-test was used for independent observations and the

Wilcoxon's signed-rank test was used for paired observations. The data cut-off date for all analyses was 1 December 2018. Microsoft Excel v16.13.1, GraphPad Prism v7.0, IBM SPSS Statistics 23, SAS v9.4 and R v3.3.2⁴⁵ were used for statistical analyses. Reported P values are unadjusted, unless stated otherwise.

Peripheral blood parameters

Baseline neutrophil, lymphocyte and eosinophil counts and LDH and C-reactive protein levels were measured according to local guidelines as part of routine diagnostics. The neutrophil-to-lymphocyte ratio was calculated as the ratio of neutrophils over lymphocytes. Baseline cytokine levels were assessed in the serum by BioLegend's LEGENDplex bead-based cytokine assay (human CD8/natural killer cell panel; lot no. 740267) according to the manufacturer's instructions.

TILs and immunohistochemistry

Formalin-fixed paraffin-embedded tissue sections were used for H&E stainings, and for CD8 (C8/144B, DAKO), PD-L1 (22C3, DAKO), CD4 (SP35, CellMarque) and FOXP3 (236A/E7, Abcam) immunohistochemistry. Immunohistochemistry of samples was performed on a BenchMark Ultra autostainer (Ventana Medical Systems). Paraffin sections of 3 μ m were deparaffinized in the instrument with EZ prep solution (Ventana Medical Systems). Heat-induced antigen retrieval was carried out using Cell Conditioning 1 (Ventana Medical Systems) for 48 min at 95 °C. Slides were counterstained with Hematoxylin and Bluing Reagent (Ventana Medical Systems). CD4 (red) and FOXP3 (DAB) were double stained. FOXP3 was detected in the first sequence (1:200 dilution, 2 h at room temperature). Bound antibody was detected using the OptiView DAB Detection Kit (Ventana Medical Systems). In the second sequence of the double-staining procedure, CD4 was detected (1:200 dilution, 1 h at room temperature) with an additional amplification step (Ventana Medical Systems). CD4 was visualized using the UltraView Universal Alkaline Phosphatase Red Detection Kit (Ventana Medical Systems). Slides were scanned at Aperio ScanScope and uploaded on Slide Score (<http://www.slidescore.com/>). Two pathologists independently evaluated the stainings digitally. The absolute CD8 count was scored manually by one pathologist. The percentage of tumor cells and sTILs was assessed by pathologists trained for TIL assessment on H&E-stained slides according to an accepted international standard from the International Immuno-Oncology Biomarker Working Group (see <http://www.tilsinbreastcancer.org/> for all guidelines on TIL assessment in solid tumors). CD8 staining was assessed on all intratumoral and stromal immune cells, whereas PD-L1 staining was assessed on both tumor cells and infiltrating immune cells separately. CD4 and FOXP3 were assessed as the percentage of the total stromal area by two pathologists.

DNA and RNA sequencing

DNA and RNA was isolated from freshly frozen sections of tissue biopsies containing at least 30% tumor cells, using the Qiagen AllPrep DNA/RNA/miRNA Universal Kit. Genomic DNA from peripheral blood cells was isolated using the QIAasympphony DSP circulating DNA kit. For exome sequencing, DNA was fragmented to 200–300-bp fragments by Covaris shearing, after which library preparation was performed using the KAPA HTP DNA Library Kit, according to the manufacturer's instructions. Exome enrichment was performed using the IDT Human Exome V1.0 Kit according to the manufacturer's instructions. Resultant libraries were sequenced with 100-bp paired-end reads on a HiSeq2500 in high-output mode using V2 chemistry (Illumina), and median sequencing depths of 146 (range: 122–217) for tumor samples and 64.7 (range: 44.6–83.2) for germline DNA samples were obtained. Raw reads were aligned to GRCh38 using the Burrows–Wheeler Aligner (bwa), followed by marking of duplicate reads by Picard MarkDuplicates. Subsequently, base quality scores were recalibrated using GATK BaseRecalibrator, and single-nucleotide variants and indels (insertions or deletions) were called using GATK MuTect⁴⁶. Variants were filtered using MuTect TLOD and NLOD with thresholds of 40 and 10, respectively, and were required to have passed all other MuTect tests (FILTER field equals 'PASS'). Variants were subsequently annotated using SnpEff 4.3t (build 2017-110-24 10:18) and variants were classified according to their most severe effect in the case of effects on multiple transcripts. Non-synonymous, exonic mutational load in coding genes was determined by summation of coding single-nucleotide variants and indels, specifically variants annotated as one of the following classes: conservative in-frame deletion, disruptive in-frame deletion, disruptive in-frame insertion, frameshift variant, missense variant, protein–protein contact, start lost, stop gained, stop lost, stop-retained variant and structural interaction variant. Copy number aberrations, discretized to integer allele-specific copy number estimates, along with purity and ploidy estimates, were obtained using the R package Sequenza (version 2.1.2)⁴⁷ with default settings. Genomic segments were identified as having undergone loss of heterozygosity if any allele (that is, the minor allele) had a copy number estimate of 0. Candidate tumor-specific neo-epitopes were determined and annotated using an in-house epitope prediction pipeline, which uses a random forest model to score the probability of surface expression of candidate neo-epitopes based

on the major prerequisites for (neo-)antigen presentation: RNA expression level (Salmon version 0.9.1)⁴⁸, proteasomal processing (NetChop version 3.1)^{49,50} and human leukocyte antigen binding (netMHCpan version 4)⁵¹. Candidate neo-epitopes that have a model prediction score lower than 0.02 are filtered out. The input variants used for the neo-epitope prediction pipeline were filtered using the default MuTect TLOD and NLOD thresholds and were required to have passed all other MuTect tests (FILTER field equals 'PASS'). Whole exome sequencing of tumor and germline DNA isolated from peripheral blood was available for 50 patients at baseline.

To obtain RNA sequencing data, strand-specific libraries were generated using the TruSeq Stranded mRNA sample preparation kit (Illumina) according to the manufacturer's instructions. The 3' end-adenylated and adapter-ligated RNA was amplified by 12 cycles of PCR. The libraries were analyzed

on a 2100 Bioanalyzer using a 7500 chip (Agilent), diluted and pooled equimolar into a multiplex sequencing pool and stored at -20°C . Resultant libraries were sequenced with 65-bp single-end reads on a HiSeq2500 in high-output mode using V4 chemistry (Illumina). Gene-specific read counts for the Ensembl version 86 build of the human transcriptome on reference genome GRCh38 were obtained by running Salmon (version 0.11.0)⁴⁸ directly on the FASTQ files using default settings, after which transcript specific read counts were collapsed to gene expression read counts using the R Bioconductor package tximport, version 1.4.0. Read counts were subsequently trimmed mean of M values (TMM)-normalized using the edgeR Bioconductor package, version 3.18.1^{52,53}. RNA sequencing data were obtained for 53 patients at baseline and 44 patients post-induction.

NanoString gene expression analysis

mRNA expression was measured with the nCounter technology, provided by NanoString Technologies. nCounter uses probes with barcodes attached to DNA oligonucleotides that directly bind to RNA. Preparation and analyses were performed according to the manufacturer's protocol using The PanCancer IO 360 gene expression panel that includes 770 genes (for research use only and not for use in diagnostic procedures). Signatures were defined as described previously^{22,23,26}. Normalization was performed by correcting for the expression of technical controls and 30 housekeeping genes included in the panel. A PAM50 spike-in panel of 30 genes was used to determine PAM50 subtypes. nCounter gene expression data were obtained for 51 patients at baseline, 45 patients post-induction and 30 patients on nivolumab.

TCR sequencing

The ImmunoSEQ Assay (Adaptive Biotechnologies) covering the CDR3 region of the human TCR β -chain was performed on DNA isolated from baseline, post-induction and on-nivolumab tumor samples. For a subset of patients, DNA was isolated from peripheral blood mononuclear cells with the Qiagen DNeasy Blood & Tissue Kit. Extracted genomic DNA was amplified in a bias-controlled multiplex PCR, followed by high-throughput sequencing. Sequences were collapsed and filtered to identify and quantitate the absolute abundance of unique TCR- β CDR3 region for further analysis. TCR sequencing data of tumor-infiltrating T cells were obtained for 48 patients at baseline, 43 patients post-induction and 29 patients on nivolumab. TCR sequencing data of peripheral blood T cells were obtained for 20 patients at baseline, post-induction and on nivolumab. The following T cell repertoire summary statistics were extracted from the Adaptive ImmunoSeq Analyzer: clonality, number of unique clones (repertoire diversity), as estimated by the Efron–Thisted estimator⁵⁴, and T cell infiltration, as measured by the fraction of T cells over nucleated cells.

Gene Set Enrichment Analysis on RNA sequencing data

To analyze which cellular processes were most strongly affected by the four induction treatments, a GSEA⁵⁵ was performed on the 50 hallmark gene sets²⁴ and separately on 4 MDSC-associated²⁵ and CD4 T cell-associated²⁷ gene sets using the flexgsea-r R package (<https://github.com/NKI-CCB/flexgsea-r>) on the TMM-normalized read counts as detailed above. Having defined a custom gene-ranking function, genes were ranked according to the P values of a pairwise Wilcoxon rank-sum test, as implemented by the `wilcox.test()` function in R. Specifically, the following gene-ranking value was used: $r(g) = \text{sign}(\text{FCg})(1 - P_g)$, in which the `sign` function returns either 1 or -1 depending on the sign of its operand, FCg reflects the median fold change (FC) between the two compared time points and P_g represents the P value of the Wilcoxon rank-sum test. During permutation steps ($n = 1,000$), samples from both time points were assigned randomly to time point and patient combinations.

PAM50 subtyping on RNA sequencing data

PAM50 subtyping was done on TMM-normalized RNA sequencing data using the `genefu` package in R, version 2.11.2⁵⁶.

Bayesian hierarchical modelling of gene expression FCs

We noticed differences between the induction treatments in the FCs between the baseline and post-induction or on nivolumab timepoints of the 12 NanoString gene set scores, related to inflammation and T-cell activation. Thus, we wanted to quantify to what degree induction cohorts were enriched for high or low FCs. As the gene expression scores for these gene sets were highly correlated, we first summarized them by taking the median (which we refer to as the ‘inflammation score’) per patient and time point. We then modelled the FCs in inflammation scores over time between the baseline (biopsy one) and post-induction (biopsy two) time points using a hierarchical Bayesian regression model. This model regularizes the effects ascribed to the induction treatments by partially pooling effects across induction arms, which increases inferential robustness. Specifically, the means of observed FCs for each induction treatment, μ_{arm} , were assumed to originate from a normal distribution, $N(\mu, \sigma_{arm})$, for which both the mean (μ) and the standard deviation (σ_{arm}) parameters were estimated using scaled Student’s t -distributions ($t(df, m, s)$) as their priors, where df denotes the Student- t ’s degrees of freedom, m represents the location of the mode and s represents the scaling to be applied to the data beforehand. We employed $df = 3$, $m = 0$ and $s = 10$, throughout, to get weakly informative priors centered at 0. Next, the observed FCs were modelled as generated by induction arm-specific normal distributions with mean μ_{arm} and standard deviation σ_{FC} , the latter of which is shared between induction arms (variation in observed FCs within arms appeared equal). Combined, this gives the following set of expressions (as graphically represented in Figure

S3.8a):

$$\begin{aligned}
 \mu &\sim t(3, 0, 10) \\
 \sigma_{arm} &\sim t(3, 0, 10) \\
 \mu_{arm} &\sim N(\mu, \sigma_{arm}) \\
 \mu_{FC} &= \mu_{arm} \\
 \sigma_{FC} &\sim t(3, 0, 10) \\
 FC &\sim N(\mu_{FC}, \sigma_{FC})
 \end{aligned}$$

In which μ_{FC} , the expected FC for an individual observation, equals μ_{arm} in this basic version of the model, but will shortly be augmented with additional co-variates. After fitting the model, we normalized the μ_{arm} -estimates of the induction treatments to that of the no induction cohort by computing and reporting the pairwise fold differences in μ_{arm} as compared to μ_{arm} of the no induction cohort (Figure S3.8b). Obtained results were robust to varying df for both σ_{arm} and σ_{FC} between 1 and 6 (data not shown).

As we noticed the inflammation score at baseline (S_{BL}) to negatively associate with the observed inflammation score FCs (Figure S3.8c), we also investigated an extension of this model in which S_{BL} influences the observed FC in a global, arm-unspecific manner by augmenting μ_{FC} :

$$\begin{aligned}
 b &\sim t(3, 0, 10) \\
 \mu_{FC} &= \mu_{arm} + bS_{BL}
 \end{aligned}$$

where all statistical definitions of the previous model, except for the superseded μ_{FC} , still apply. Second, we tested whether describing the effect of having a clinical response to nivolumab would abrogate the intercohort differences, as we did observe higher FCs in responding patients when comparing on-nivolumab (biopsy three) and baseline (biopsy one) time points, which is not surprising considering the way these gene sets were selected (Figure 3.3g). Similarly to the previous expansion of μ_{FC} with $b S_{BL}$, we thus augmented μ_{FC} with $r R$, in which r describes the effect attributed to having a clinical response

(modeled as $cr \sim t(3, 0, 10)$) and R is an indicator variable for clinical response to nivolumab (Figure S3.8d). Third, we were interested in testing whether having previous treatment for metastatic disease before enrollment in the TONIC trial affected the observed upregulation. This was motivated by the fact that we observed a trend towards a higher clinical response rate in patients with no previous lines of treatment than in patients with one or more lines of previous treatment (Figure S3.5c).

Thus, we further expanded the expression for μ_{FC} with lL , in which l describes the effect attributed to having multiple treatment lines (prioritized as $l \sim t(3,0,10)$) and L is an indicator variable representing whether palliative treatment was administered (Figure S3.8e). Finally, we also tested the relationship between having metastases restricted to the lymph nodes as opposed to other organs (Figure S3.8f), expanding the expression for μ_{FC} with nN , in which n describes the effect attributed to having metastases restricted to lymph nodes (prioritized as $n \sim t(3,0,10)$) and N is an indicator variable for having lymph node-restricted metastases.

Testing various combinations of the four extra covariates described in the previous paragraph revealed that the inclusion of extra covariates minimally influenced the coefficients assigned to other covariates (Figure S3.8h). The exception to this is r , which was reduced by about fourfold with the inclusion of other covariates. The full model, including all of the extra covariates, shows that the baseline inflammation score (SBL) and lymph node-restricted metastases were most strongly associated with FCs in the inflammation score, besides the differential FCs apparently induced by the tested induction treatments.

These models were evaluated using the probabilistic programming language Stan⁵⁷, interfaced in R using the R package rstan (version 2.17.3). Ten chains of no-U-turn-sampler Markov chain Monte Carlo (MCMC) simulations were run for 100,000 iterations, of which 25,000 served as warm-up iterations. Sampling convergence was sufficient for all models as Rhat values were all 1. Inter-arm comparisons between μ_{arm} and μ_{arm}' were performed by extracting parameter values from non-warm-up MCMC iterations (using rstan::extract) for both arms and computing the proportion of iterations for which μ_{arm} equaled or exceeded μ_{arm} .

The stan program is available on request.

BRCA1-like classification based on copy number profiles

A BRCA1 classifier originally had been trained using the nearest shrunken centroids algorithm on bacterial artificial chromosome (BAC) array comparative genomic hybridization data⁵⁸. Data from platforms of higher resolution can be used to obtain reliable BRCA1-like classification⁵⁹. In this study, GC-content-corrected allele imbalance log ratios, to be used for downstream copy number estimates, were obtained from whole-exome sequencing data using the Sequenza R package⁴⁷. To apply the BRCA1-like classifier, these estimates had to be preprocessed to comply with the format of the original training set. LiftOver was used to map the genomic locations from GRCh38 to hg19, the reference genome on which the BRCA1-like classifier was validated. Average log ratios for each of the original 3,277 BAC-array segments were computed by averaging the binned log ratios within 500 kb upstream and downstream of the central genomic position of the BAC clone. Missing values due to a lack of coverage were subsequently replaced using linear interpolation between adjacent features on the same chromosome. On average, 487 probes were estimated per sample, of which on average 372 had directly surrounding probes available for interpolation. The mean and maximum genomic

distances between estimated and nearest measured segments were 2 Mb (2 segments) and 7 Mb (7 segments), respectively. The distribution of resulting whole-exome sequencing (WES)-derived segment log ratios differed in the mean from that of previously obtained BAC-derived segment log ratios of patients with TNBC⁶⁰. To correct the WES segments, we first fitted a linear model (iteratively reweighted least squares) between the sorted segment-wise averages of the WES and BAC segments. The WES data were then corrected using the following expression $f_c = \alpha + \beta f_o$, in which f_o and f_c represent the original and corrected segments, respectively, and α (0.16) and β (0.97) represent the fitted parameters. This yielded highly similar distributions between the newly obtained WES-derived and original BAC-derived log ratio estimates (Pearson's $r^2 = 0.96$), but the former remained slightly right skewed. Finally, the WES data were classified with the established nearest shrunken centroid classifier, using a previously established value of at least 0.63 to be classified as BRCA1-like (as used in earlier work⁶⁰; <http://ccb.nki.nl/software/nkibrca/>).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

DNA and RNA sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAS0001003535 and will be made available from the corresponding author on reasonable request. Data requests will be reviewed by the institutional review board of the NKI and applying researchers will need to sign a data access agreement with the NKI after approval. The TCR sequencing data are available from Adaptive Biotechnologies, but restrictions apply to their availability. However, data are available from the corresponding author on reasonable request and with permission of Adaptive Biotechnologies.

Supplemental Figures

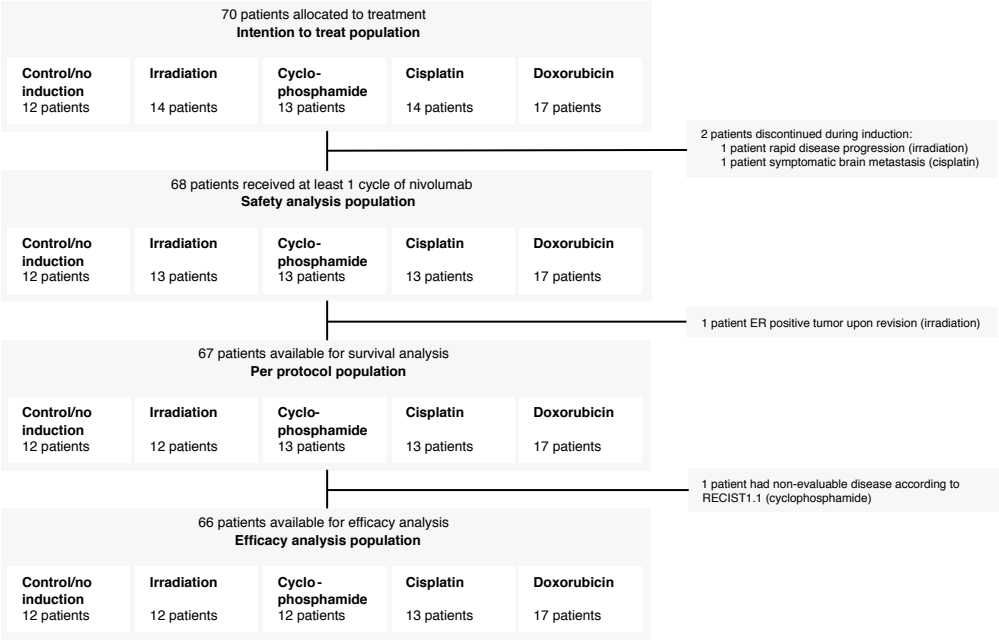


Figure S3.1: CONSORT diagram. Flowchart for the allocation of subjects enrolled in the trial.

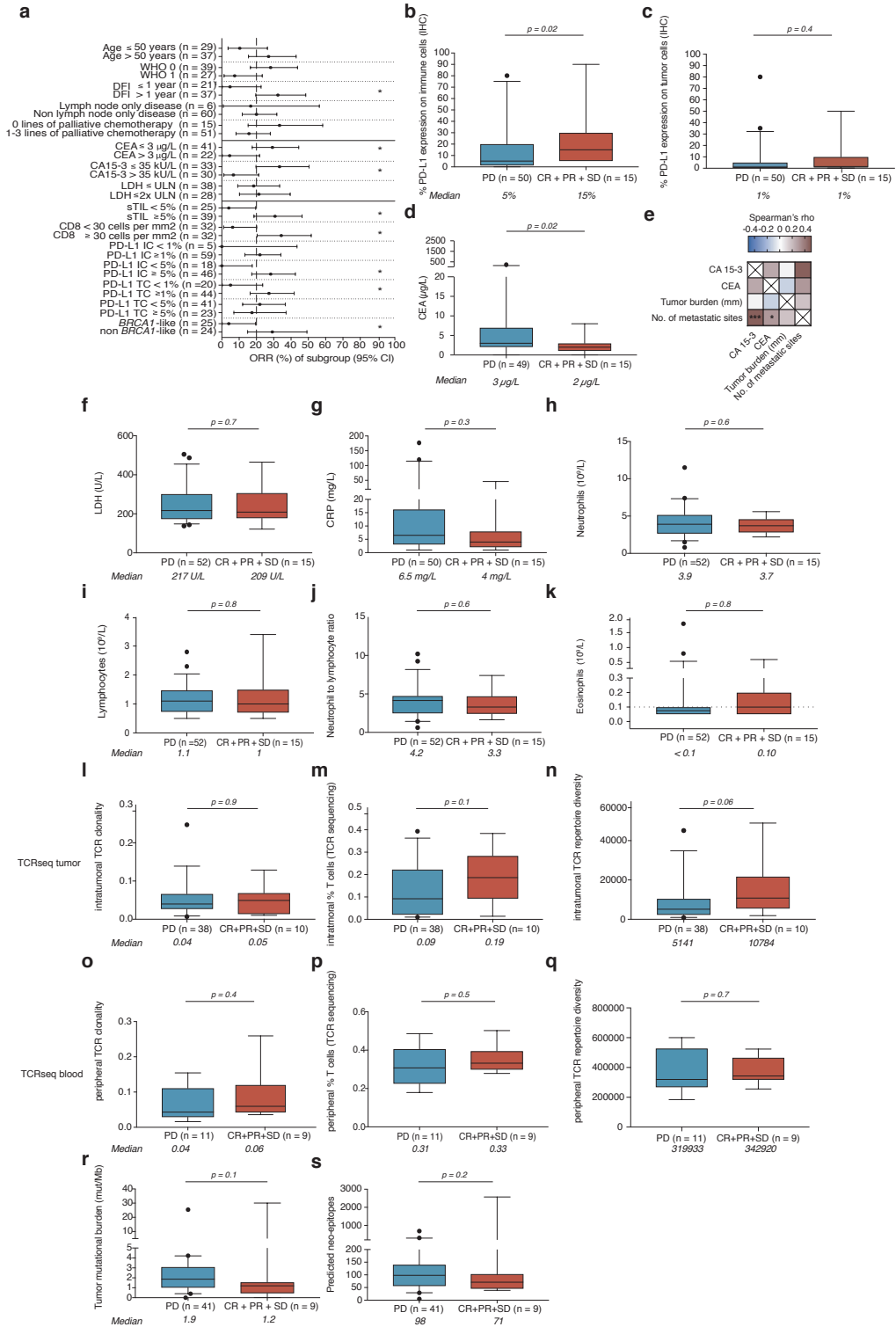


Figure S3.2: Clinical and other baseline parameters associated with response. Boxplots represent the median, 25th and 75th percentiles and the vertical bars span the 5th and 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney *U*-test (unadjusted *P* values).

a ORR per subgroup. Depicted is the ORR (CR + PR of *n* = 66) per subgroup. Cut-offs are set at the median for carcinoembryonic antigen (CEA), CA 15-3, sTIL and CD8. Statistical significance was determined by a two-sided Fisher's exact test. **P* < 0.05. WHO, WHO performance status. ¹Patients with de novo metastatic disease at diagnosis were excluded (*n* = 8).

b PD-L1 expression on immune cells.

c PD-L1 expression on tumor cells.

d Serum levels of CEA.

e Correlation of CA 15-3 and CEA with tumor burden and the number of metastatic sites. Spearman correlation coefficients are depicted. Tumor burden was measured as the sum of all target lesions in millimeters; **P* < 0.05; ****P* < 0.001.

f LDH levels.

g C-reactive protein (CRP) levels.

h Neutrophil counts.

i Lymphocyte counts.

j Neutrophil-to-lymphocyte ratio (NLR).

k Eosinophil counts. The dashed line indicates the detection limit.

l Intratumoral TCR clonality.

m Percentage of intratumoral T cells by TCR sequencing.

n Intratumoral TCR repertoire diversity.

o TCR clonality in the peripheral blood.

p Percentage of T cells by TCR sequencing in the peripheral blood.

q TCR repertoire diversity in the peripheral blood.

r Non-synonymous tumor mutational burden (TMB).

s Predicted neo-epitopes.

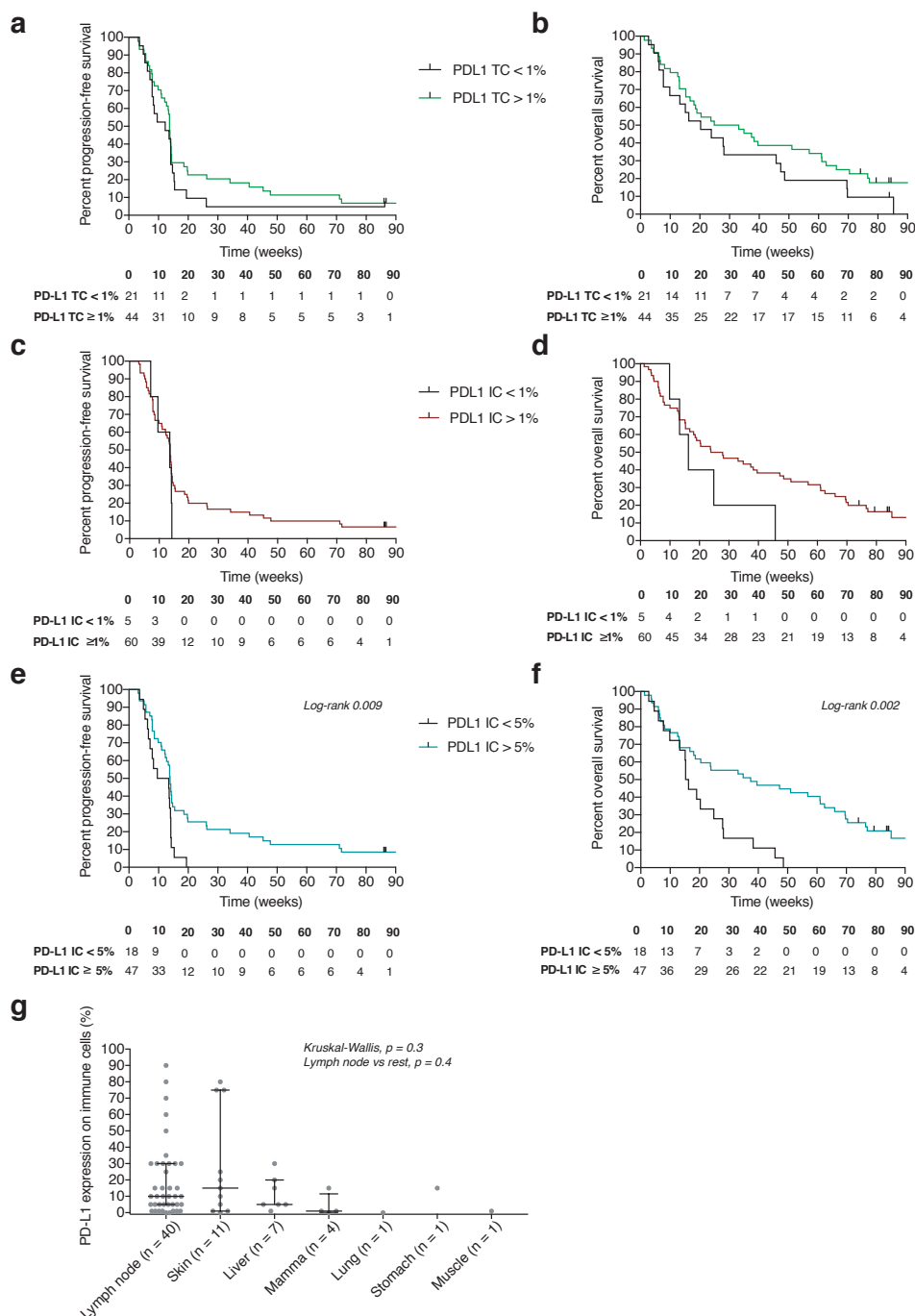


Figure S3.3: Extended Data Figure 3. Baseline PD-L1 expression.

a PFS and PD-L1 expression on tumor cells. The Kaplan-Meier curve displays the proportion of patients free of progression, stratified by PD-L1 expression on tumor cells. A cut-off of 1% is used. The table lists the number of patients at risk.

b Overall survival and PD-L1 expression on tumor cells. The Kaplan-Meier curve displays overall survival,

stratified by PD-L1 expression on tumor cells. A cut-off of 1% is used.

c PFS and PD-L1 expression on tumor-infiltrating immune cells. The Kaplan–Meier curve displays the proportion of patients free of progression, stratified by PD-L1 expression on tumor-infiltrating immune cells. A cut-off of 1% is used.

d Overall survival and PD-L1 expression on tumor-infiltrating immune cells. The Kaplan–Meier curve displays overall survival, stratified by PD-L1 expression on tumor-infiltrating immune cells. A cut-off of 1% is used.

e PFS and PD-L1 expression on tumor-infiltrating immune cells. The Kaplan–Meier curve displays the proportion of patients free of progression, stratified by PD-L1 expression on tumor-infiltrating immune cells. A cut-off of 5% is used.

f Overall survival and PD-L1 expression on tumor-infiltrating immune cells. The Kaplan–Meier curve displays overall survival, stratified by PD-L1 expression on tumor-infiltrating immune cells. A cut-off of 5% is used.

g PD-L1 expression on tumor-infiltrating immune cells and site of metastasis. PD-L1 expression per biopsy site at baseline is shown. Dots reflect the medians and whiskers reflect the interquartile ranges. IC, tumor-infiltrating immune cells; TC, tumor cells.

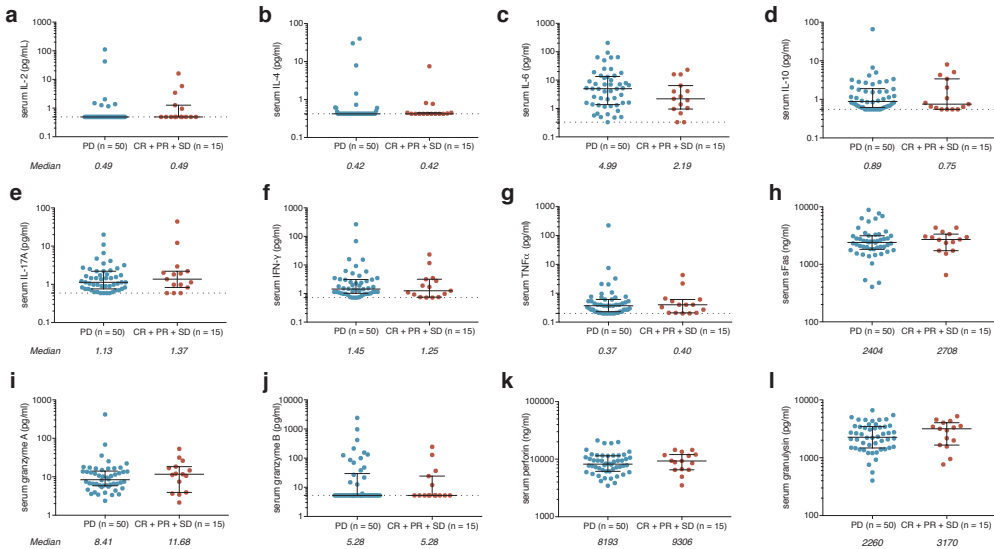


Figure S3.4: Extended Data Figure 4. Baseline serum cytokine levels. Cytokine levels were determined by a validated bead-based assay. Dots and whiskers represent medians and interquartile ranges, respectively. The dashed lines indicate the detection limit.

a IL-2 levels. IL-2 levels were detectable in five patients with clinical benefit and seven patients with progressive disease.

b IL-4 levels. IL-4 levels were detectable in 5 patients with clinical benefit and 11 patients with progressive disease.

c IL-6 levels. IL-6 levels were detectable in 13 patients with clinical benefit and 49 patients with progressive disease.

d IL-10 levels. IL-10 levels were detectable in 11 patients with clinical benefit and 41 patients with progressive disease.

e IL-17A levels. IL-17A levels were detectable in 12 patients with clinical benefit and 46 patients with progressive disease.

f IFN-γ levels. IFN-γ levels were detectable in 13 patients with clinical benefit and 47 patients with progressive disease.

g TNF-α levels. TNF-α levels were detectable in 11 patients with clinical benefit and 45 patients with progres-

sive disease.

h Soluble Fas (sFas) levels. sFas levels were detectable in all tested patients.

i Granzyme A levels. Granzyme A levels were detectable in all tested patients.

j Granzyme B levels. Granzyme B levels were detectable in 5 patients with clinical benefit and 17 patients with progressive disease.

k Perforin levels. Perforin levels were detectable in all tested patients.

l Granulysin levels. Granulysin levels were detectable in all tested patients.

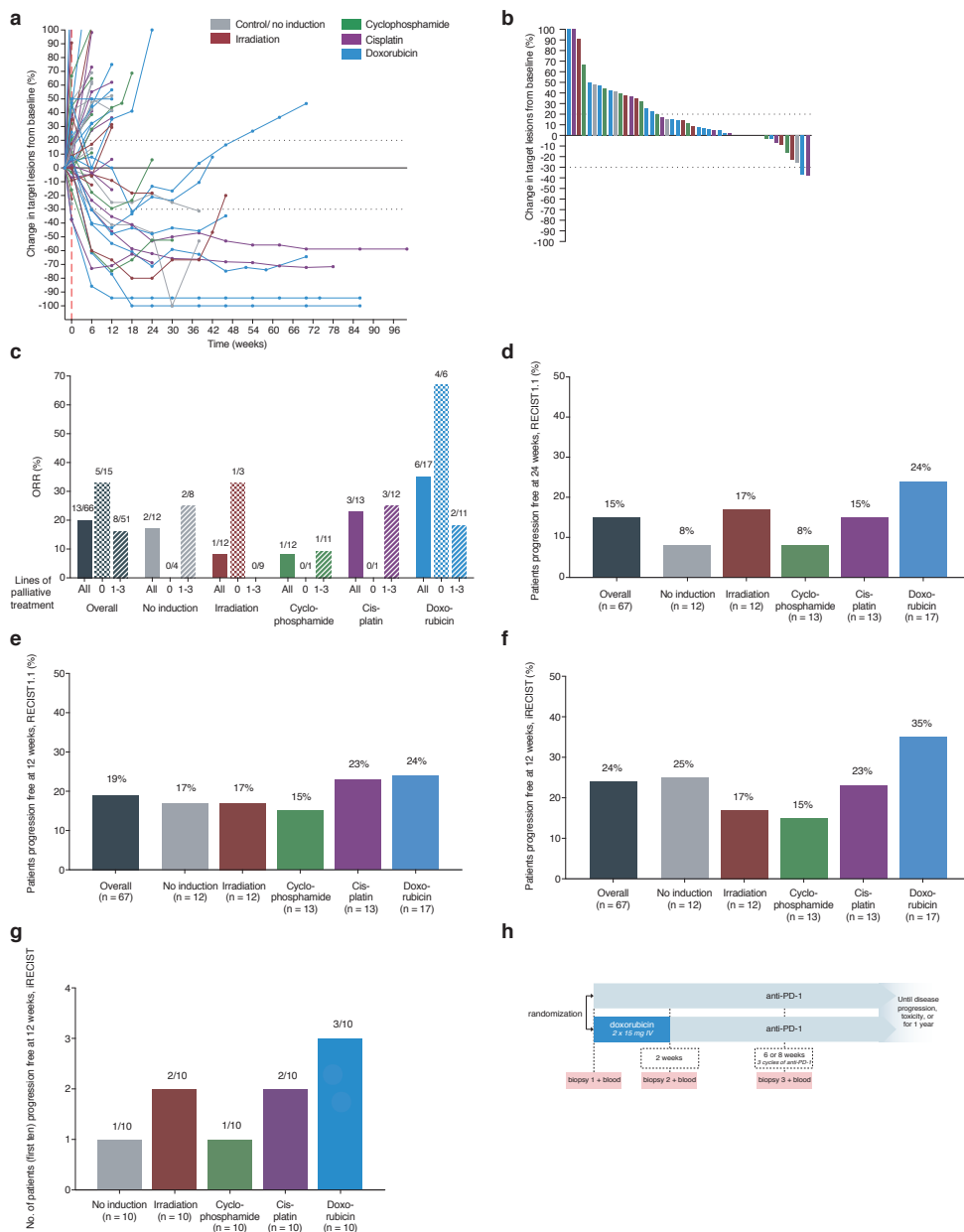


Figure S3.5: Extended Data Figure 5. Anti-tumor activity of nivolumab

a Changes in target lesions over time, reflecting the depth and duration of response. Every line represents one patient, and every dot is one time point. The colors reflect induction treatment. The y axis was cut-off at 100% for illustration purposes. Dotted black lines indicate the response as described by RECIST1.1.

b Waterfall plot depicting the change in target lesions from baseline to post-induction. Every bar represents one patient and the colors correspond to induction treatment. The y axis was cut-off at 100% for illustration purposes. Dotted black lines indicate the response as described by RECIST1.1.

c ORR per cohort and according to lines of palliative treatment. The bars with no pattern depict the overall response rate in all patients, the bars with a dotted pattern depict the overall response rate in first-line-treated patients and the bars with a lined pattern depict the overall response rate in the second-to-fourth-line-treated patients. The numbers above the bars reflect the number of responding patients (CR + PR) over the total number of patients in that subgroup.

d Proportion of patients free of progression at 24 weeks. Measured from randomization according to RECIST1.1 (primary end point).

e Proportion of patients free of progression at 12 weeks. Measured from nivolumab initiation (including response evaluation performed at 14 weeks from randomization) according to RECIST1.1.

f Proportion of patients free of progression at 12 weeks. Measured from nivolumab initiation (including response evaluation performed at 14 weeks from randomization) according to iRECIST.

g Number of patients free of progression at 12 weeks in the first 10 included patients. Measured from nivolumab initiation (including response evaluation performed at 14 weeks from randomization) according to iRECIST.

h Trial design of TONIC stage 2. Patients are randomized between (1) induction treatment of 2 weeks with doxorubicin followed by anti-PD-1 or (2) start with anti-PD-1 without induction treatment.

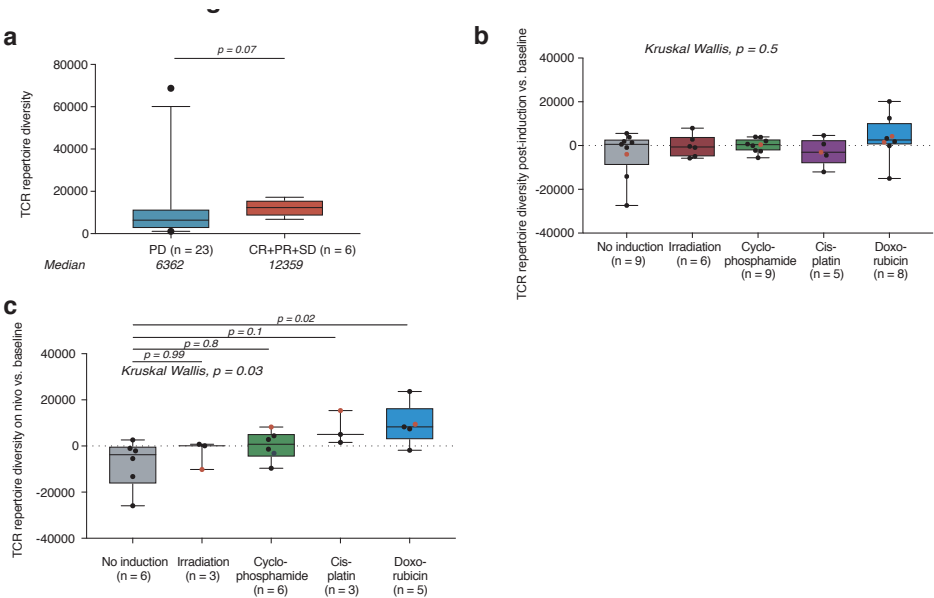


Figure S3.6: TCR repertoire diversity during an ongoing anticancer response on nivolumab and changes observed after induction treatments

a TCR repertoire diversity on nivolumab treatment (biopsy three). TCR repertoire size was estimated using the Efron–Thisted method⁵⁴ and represents the number of unique intratumoral clones. The boxes in boxplots represent the median and interquartile ranges and the whiskers represent the 5th and 95th percentiles.

b Fold change (FC) in the number of unique intratumoral TCR clones (TCR repertoire diversity) after induction

treatment versus baseline (biopsy two versus biopsy one). Every dot represents one patient. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change. TCR repertoire size was estimated using the Efron–Thisted method⁵⁴. The boxes in the boxplots represent the median and interquartile ranges and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups.

c FC in the number of unique intratumoral TCR clones (TCR repertoire diversity) after nivolumab treatment versus baseline (biopsy three versus biopsy one). Every dot represents one patient. Patients with clinical benefit are highlighted with a red dot. The dotted black line indicates no change. TCR repertoire size was computed using the Efron–Thisted method⁵⁴. The boxes in the boxplots represent the median and interquartile ranges and the whiskers represent the full range. Statistical significance was tested with a Kruskal–Wallis test for all groups followed by Dunn’s tests between the induction treatment groups and the control group (P values are adjusted).

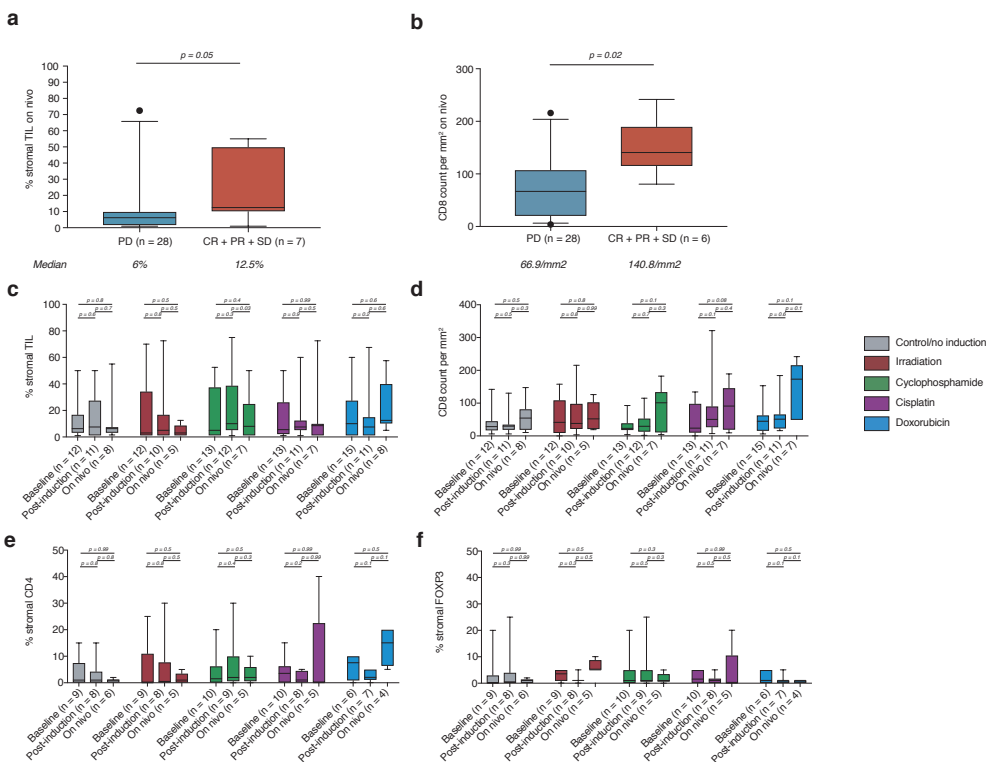


Figure S3.7: Histological characteristics of tumors during an ongoing anticancer response on nivolumab and changes observed after induction treatments

a sTILs in on-nivolumab biopsies (biopsy three), as determined according to guidelines of the TIL working group on a H&E staining. The boxes in the boxplots represent the median and interquartile ranges, and the whiskers represent the 5th and 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney U -test (unadjusted P value).

b CD8 cell count per mm² in on-nivolumab biopsies (biopsy three). The boxes in the boxplots represent medians with interquartile ranges, and the whiskers span the 5th to 95th percentiles. Statistical significance was tested with a two-tailed Mann–Whitney U -test (unadjusted P value).

c sTILs per cohort. The boxes in the boxplots represent medians with interquartile ranges, and the whiskers span the 5th to 95th percentiles. Statistical significance was tested on paired biopsies with the Wilcoxon signed-rank

test (two-tailed and unadjusted *P* value).

d CD8 cell count per mm² per cohort. The boxes in the boxplots represent medians with interquartile ranges, and the whiskers span the 5th to 95th percentiles. Statistical significance was tested on paired biopsies with the Wilcoxon signed-rank test (two-tailed and unadjusted *P* value).

e Stromal CD4 per cohort. The percentage of CD4 of the total stromal area was assessed. The boxes in the boxplots represent medians with interquartile ranges, and the whiskers span the 5th to 95th percentiles. Statistical significance was tested on paired biopsies with the Wilcoxon signed-rank test (two-tailed and unadjusted *P* value).

f Stromal FOXP3 per cohort. The percentage of FOXP3 of the total stromal area was assessed. The boxes in the boxplots represent medians with interquartile ranges, and the whiskers span the 5th to 95th percentiles. Statistical significance was tested on paired biopsies with the Wilcoxon signed-rank test (two-tailed and unadjusted *P* value).

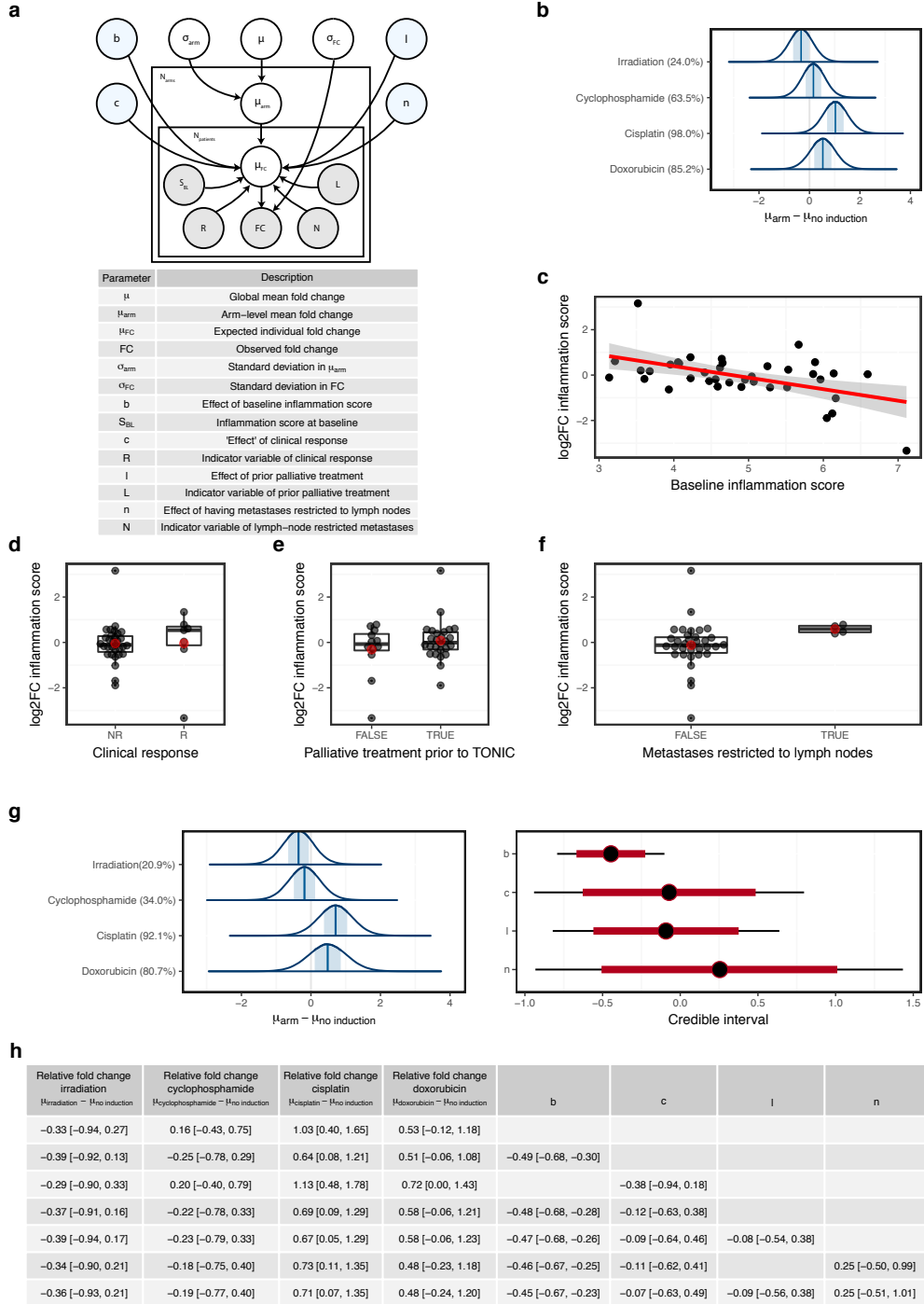


Figure S3.8: Bayesian hierarchical regression analysis of inflammation-related gene set FCs to investigate differences in upregulation between induction arms.
a Plate model representation of the hierarchical model describing the FCs between baseline and post-induction.

White-colored variables are inferred from the data using the model, and blue-colored variables are incorporated in extensions of the basic model. The boxes reflect repetition of the variables, $N_{\text{arms}} = 5$ and N_{patients} varies between arms. Data were available for 38 patients.

b Distributions of posterior parameter estimates for the basic hierarchical regression model. The percentages in the vertical labels represent probabilities of exceeding the control arm (the proportion of the distribution above zero).

c Effect of the baseline inflammation score on the observed FC in the inflammation score. Shown in red is the conditional mean (linear regression) with the 95% confidence interval shaded gray. The intercept of this line is not explicitly included in the model as it is already implicitly modelled by the μ .

d Association between clinical response and the observed FC in the inflammation score. Red dots indicate the means. The boxes in the boxplots represent medians and interquartile ranges, and the whiskers span 1.5 times the interquartile range.

e Association between previous lines of palliative treatment and the observed FC in the inflammation score. Boxplots are as in **d**. **f**, Association between lymph node-only metastasis and the observed FC in the inflammation score. Boxplots are as in **d**. **g**, Distributions of posterior parameter estimates for the full hierarchical regression model including all considered covariates. Format as in **b**. The points indicate the medians, the red lines indicate the 10–90% percentiles and the black lines indicate the 2.5–97.5% percentiles.

h Robustness of coefficients with inclusion of extra covariates. Shown are the medians of the posterior parameter distributions with the 10th and 90th percentiles for 7 different models, including and excluding combinations of the non-induction arm covariates.

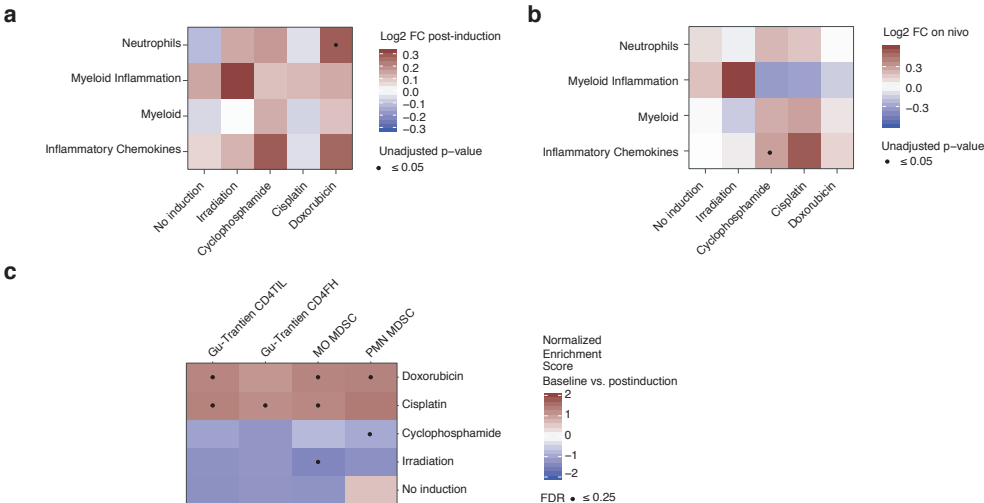


Figure S3.9: Extended Data Figure 9. Treatment-induced changes in myeloid cell-related and CD4 cell-related gene signatures.

a Heatmap of post-induction FCs in gene expression signatures^{23,26} (NanoString gene expression data) in post-induction samples (biopsy two) compared to baseline (biopsy one). Depicted is the log₂ FC in the median gene expression of paired biopsies. Statistical significance (two-sided Wilcoxon signed-rank test; unadjusted P value) is highlighted with a black dot.

b Heatmap of on-nivolumab FCs in gene expression signatures^{23,26} (NanoString gene expression data) in samples taken on nivolumab (biopsy three) compared to baseline (biopsy one). Depicted is the log₂ FC in the median gene expression of paired biopsies. Statistical significance (two-sided Wilcoxon signed-rank test; unadjusted P value) is highlighted with a black dot.

c GSEA of selected gene sets related to myeloid cells and CD4 T cells^{25,27}, performed on whole-transcriptome

RNA sequencing of pre-induction and post-induction samples (biopsy two compared to biopsy one). Cells are colored according to normalized enrichment scores, and Benjamini–Hochberg (false discovery rate (FDR))-corrected *P* values equaling or below 0.25 are highlighted with black dots. CD4FH, follicular helper CD4 T cells; CD4TIL, tumor-infiltrating CD4 T cells; MO MDSC, monocytic MDSC; PMN MDSC, polymorphonuclear MDSC.

Supplementary Note

Eligibility criteria:

- Metastatic or incurable locally advanced triple negative breast cancer with confirmation of ER and HER2 negativity (ER <10% and HER2 0,1 or 2 in the absence of amplification as determined by in situ hybridization) on a histological biopsy of a metastatic lesion
- Metastatic lesion accessible for histological biopsy (Mandatory biopsies: pre-induction treatment, post-induction treatment, 6-weeks. Optional biopsies: 12-weeks, at progression, of irradiated site). The pre-induction treatment biopsy has to contain sufficient tumor content (≥ 100 tumor cells); subjects with samples that have insufficient tumor content will require re-biopsy prior to induction treatment. Interval between last systemic treatment and pre-induction biopsy has to be at least 14 days
- 18 years or older
- Maximum of three lines of chemotherapy for metastatic disease and with evidence of progression of disease
- Measurable or evaluable disease according to RECIST 1.1
- Metastatic lesion accessible for radiation with 1x20Gy or 3x8Gy
- No uncontrolled intercurrent illness including, but not limited to, ongoing or active infection, symptomatic congestive heart failure, unstable angina pectoris
- Subjects with brain metastases are eligible if these are not symptomatic. Subjects who received prior treatment for brain metastases should be free of progression on magnetic resonance imaging (MRI) for at least 4 weeks after treatment is completed and prior to first dose of study drug administration. There must also be no requirement for immunosuppressive doses of systemic corticosteroids (> 10 mg/day prednisone equivalents) for at least 2 weeks prior to study drug administration
- No known history of leptomeningeal disease localization
- No history of having received other anticancer therapies within 2 weeks of start of the study drug

-
- No history of immunodeficiency, autoimmune disease, conditions requiring immunosuppression (>10 mg daily prednisone equivalents) or chronic infections. Subjects with vitiligo, diabetes mellitus type I, psoriasis not requiring systemic treatment or resolved childhood asthma/atopy would be an exception to this rule. Subjects that require intermittent use of bronchodilators, inhaled steroids, or local steroid injections would not be excluded from the study. Subjects with hypothyroidism stable on hormone replacement, Sjögren's syndrome or conditions not expected to recur in the absence of an external trigger will not be excluded from the study. Adrenal replacement doses >10 mg daily prednisone equivalents are permitted in the absence of active autoimmune disease
 - No prior treatment with an anti-PD-1, anti-PD-L1, anti-PD-L2, or anti-CTLA-4 antibody
 - No live vaccine within 30 days of planned start of study therapy.
 - No active other cancer.
 - No positive test for hepatitis B surface virus surface antigen (HBV sAg) or hepatitis C virus ribonucleic acid (HCV antibody) indicating acute or chronic infection
 - WHO performance status of 0 or 1 (Appendix B)
 - No history of uncontrolled serious medical or psychiatric illness
 - Absence of any psychological, familial, sociological or geographical condition potentially hampering compliance with the study protocol and follow-up schedule
 - No current pregnancy or breastfeeding. Women of childbearing potential (WOCBP*) must use adequate contraceptive protection. WOCBP should use an adequate method to avoid pregnancy for 23 weeks (30 days plus the time required for nivolumab to undergo five half-lives) after the last dose of investigational drug. WOCBP must have a negative serum or urine pregnancy test
 - WBC $\geq 2.0 \times 10^9/L$, ANC $\geq 1.5 \times 10^9/L$, platelets $\geq 100 \times 10^9/L$, Hemoglobin ≥ 5.0
 - Bilirubin $< 1.5 \times$ upper limit of the normal range (ULN), except subjects with Gilbert Syndrome; alkaline phosphatase $< 2.5 \times$ ULN ($< 5 \times$ ULN in case of liver metastases, and $< 7 \times$ ULN in case of bone metastases); transaminases (ASAT/ALAT) $< 3 \times$ ULN (and $< 5 \times$ ULN in case of liver metastases), LDH $< 2 \times$ ULN
 - Calculated (Cockcroft-Gault) or measured creatinine clearance > 40 mL/min
 - Signed written informed consent

*) Women of childbearing potential” is defined as any female who has experienced menarche and who has not undergone surgical sterilization (hysterectomy or bilateral oophorectomy) or who is not postmenopausal. Menopause is defined clinically as 12 months of amenorrhea in a woman over 45 in the absence of other biological or physiological causes.

Table S1: Baseline characteristics of the intention-to-treat population per cohort. Clinical baseline characteristics of all allocated patients separated by induction treatment cohort ($n = 70$).

	Cohort 1, Control/no induction (n = 12)	Cohort 2, Irradiation (n = 14)	Cohort 3, Cyclophos- phamide (n = 13)	Cohort 4, Cisplatin (n = 14)	Cohort 5, Doxoru- bicin (n = 17)
Median age, years (range)	53 (33-69)	47 (33-68)	51 (35-68)	53 (41-70)	46 (29-68)
Female, n (%)	12 (100)	14 (100)	13 (100)	14 (100)	17 (100)
WHO performance status, n (%)					
0	7 (58)	6 (43)	6 (46)	8 (57)	14 (82)
1	5 (42)	8 (57)	7 (54)	6 (43)	3 (18)
gBRCA1/2, n (%)					
Mutation	3 (25)	0 (0)	0 (0)	1 (7)	2 (12)
Wildtype	7 (58)	12 (86)	11 (85)	8 (57)	12 (71)
Unknown	2 (17)	2 (14)	2 (15)	5 (36)	3 (18)
Location of metas- tasis, n (%)					
Lymph node only	1 (8)	1 (7)	0 (0)	0 (0)	4 (24)
Visceral metasta- sis	7 (58)	12 (86)	11 (85)	10 (71)	10 (59)
Other metastasis	4 (33)	1 (7)	2 (15)	4 (29)	3 (18)
No. of prior ther- apies for metastatic disease, n (%)					
0	4 (33)	4 (29)	2 (15)	1 (7)	6 (35)
1	6 (50)	7 (50)	6 (46)	7 (50)	8 (47)
2-3	2 (17)	3 (21)	5 (38)	6 (43)	3 (18)
Previous neoadju- vant or adjuvant therapy, n (%)	11 (92)	13 (93)	11 (85)	10 (71)	14 (82)
Previous chemotherapy exposure, n (%)					
Taxane	11 (92)	13 (92)	12 (93)	12 (86)	16 (94)
Anthracycline	11 (92)	14 (100)	11 (85)	10 (71)	14 (82)
Platinum	12 (100)	7 (50)	5 (38)	8 (57)	10 (59)
Capecitabine	6 (50)	6 (43)	8 (62)	8 (57)	6 (35)
Disease free inter- val, n (%)					

De novo metastatic disease	1 (8)	1 (7)	2 (15)	3 (21)	2 (12)
Disease free interval ≤ 12 months	4 (33)	8 (57)	5 (38)	2 (14)	4 (24)
Disease free interval > 12 months	7 (58)	5 (36)	6 (46)	9 (64)	11 (65)
LDH level, n (%)					
≤ ULN	7 (58)	7 (50)	8 (62)	6 (43)	11 (65)
≤ 2x ULN	5 (42)	7 (50)	5 (38)	8 (57)	6 (35)
Stromal tumor infiltrating lymphocytes, n (%)					
Not available	0 (0)	2 (14)	0 (0)	1 (7)	2 (12)
< 5%	4 (33)	7 (50)	4 (31)	4 (29)	6 (35)
≥ 5%	8 (67)	5 (36)	9 (69)	9 (64)	9 (53)
PD-L1 expression on tumor cells, n (%)					
Not available	0 (0)	2 (14)	1 (8)	1 (7)	1 (6)
≥ 1% on tumor cells	8 (67)	10 (71)	9 (69)	8 (57)	9 (53)
≥ 5% on tumor cells	3 (25)	7 (50)	5 (38)	4 (29)	4 (24)
PD-L1 expression on immune cells, n (%)					
Not available	0 (0)	2 (14)	1 (8)	1 (7)	1 (6)
≥ 1% on immune cells	11 (92)	10 (71)	12 (92)	12 (86)	15 (88)
≥ 5% on immune cells	9 (75)	8 (57)	8 (62)	11 (79)	11 (65)
No. of nivolumab cycles, median (range)	5 (2-20)	2 (0-24)	5 (1-16)	6 (0-42)	6 (1-37)

Table S2: Baseline characteristics of the efficacy analysis population. Clinical baseline characteristics of the patients included in the efficacy analysis ($n = 66$)

Efficacy analysis population ($n = 66$)		
Median age, years (range)	51 (29-70)	
	<i>n</i>	%
WHO performance status		

0	39	59%
1	27	41%
gBRCA1/2		
Mutation	6	9%
Wildtype	47	71%
Unknown	13	20%
Location of metastasis		
Lymph node only	6	9%
Visceral metastasis	47	71%
Other metastasis	13	20%
No. of prior therapies for metastatic disease		
0	15	23%
1	33	50%
2-3	18	27%
Previous neoadjuvant or adjuvant therapy	56	85%
Previous chemotherapy exposure		
Taxane	60	91%
Anthracycline	56	85%
Platinum	38	58%
Capecitabine	33	50%
Disease Free Interval		
De novo metastatic disease	8	12%
Disease Free Interval \leq 12 months	21	32%
Disease Free Interval $>$ 12 months	37	56%
LDH level		
\leq ULN	38	58%
\leq 2x ULN	28	42%
PD-L1 expression on tumor cells		
Not available	2	3%
\geq 1% on tumor cells	41	62%
\geq 5% on tumor cells	23	35%
PD-L1 expression on immune cells		
Not available	2	3%
\geq 1% on immune cells	59	89%
\geq 5% on immune cells	46	70%

¹Two patients with clear evidence of progression died on study. One patient died from bacterial peritonitis probably due to progressive intestinal metastasis. One patient with progressive pleural fluid and lymphangitic carcinomatosis died from acute respiratory insufficiency not otherwise specified while being hospitalized.

Table S3: Maximum grade of nivolumab or induction treatment-related adverse events. Maximum grade of treatment-related adverse event, separated by induction treatment cohort. Denoted are the maximum grade adverse events during nivolumab treatment and during induction treatment.

	Any grade	Grade 3	Grade 4	Grade 5
Induction related (n = 68)	19 (28)	2 (3)	0	0
Control/no induction (n = 12)	1 (8)	0	0	0
Irradiation (n = 13)	3 (23)	0	0	0
Cyclophosphamide (n = 13)	3 (23)	1 (8)	0	0
Cisplatin (n = 13)	8 (62)	1 (8)	0	0
Doxorubicin (n = 17)	4 (24)	0	0	0
Nivolumab related (n = 68)	55 (81)	9 (13)	2 (3)	2 (3) ¹
Control/no induction (n = 12)	9 (75)	2 (17)	0	1 (8)
Irradiation (n = 13)	10 (77)	3 (23)	0	1 (8)
Cyclophosphamide (n = 13)	13 (100)	2 (15)	2 (15)	0
Cisplatin (n = 13)	11 (85)	2 (15)	0	0
Doxorubicin (n = 17)	12 (71)	0	0	0

Table S4: Nivolumab related adverse events. Nivolumab related adverse events of any grade, occurring in at least 5% of patients, and all grade 3-5 adverse events, all immune-related adverse events and immune-related events of special interest.

n = 68	Any grade, n (%)	Grade 3, n (%)	Grade 4, n (%)	Grade 5, n (%)
Fatigue	16 (24)	0	0	0
Alanine aminotransferase increased	13 (19)	0	0	0
Aspartate aminotransferase increased	13 (19)	1 (1)	0	0
Hypothyroidism	12 (18)	0	0	0
Diarrhea	9 (13)	0	0	0
Dyspnea¹	8 (12)	2 (3)	0	0
Gamma-glutamyltransferase increased	8 (12)	2 (3)	1 (1)	0
Alkaline phosphatase increased	7 (10)	1 (1)	0	0
Infusion related reaction	7 (10)	0	0	0
Fever	6 (9)	0	0	0
Flu like symptoms	5 (7)	0	0	0
Pain	5 (7)	0	0	0
Serum amylase increased	4 (6)	0	1 (1)	0
Hyperthyroidism	4 (6)	0	0	0
Sarcoidosis	4 (6)	1 (1)	0	0
Lipase increased	3 (4)	1 (1)	0	0
Anemia	2 (3)	1 (1)	0	0
Abdominal infection	1 (1)	0	0	1 (1)
Anorexia	1 (1)	1 (1)	0	0
Febrile neutropenia	1 (1)	1 (1)	0	0
Hypertension	1 (1)	1 (1)	0	0
Hyponatremia	1 (1)	1 (1)	0	0
Respiratory insufficiency	1 (1)	0	0	1 (1)

Syncope	1 (1)	1 (1)	0	0
Immune-related colitis	1 (1)	0	0	0
Immune-related gastritis	1 (1)	0	0	0
Immune-related pneumonitis	1 (1)	0	0	0

Table S5: Efficacy of nivolumab in the efficacy analysis population. Efficacy analysis was performed using the efficacy analysis population ($n = 66$) as determined by iRECIST. Complete and partial responses had to be confirmed on at least one subsequent CT scan. 1 patient was non-evaluable according to RECIST1.1, but had clinical benefit for more than 24 weeks. Progression-free survival was calculated in the per protocol population ($n = 67$) from randomization to date of progression according to RECIST1.1 or iRECIST. Eleven patients had unequivocal clinical progression before the first response evaluation. Duration of response was assessed using Kaplan Meier curves, with censoring of ongoing responses, and calculated from time of first PR or CR until progression according to iRECIST.

Objective response rate iRECIST, n, % (95% CI)	13	20% (11-31)
Clinical benefit rate iRECIST, n, % (95% CI)	14	21% (12-33)
Best overall response iRECIST, n (%)		
Complete response	2	3%
Partial response	11	17%
Stable disease > 24 weeks	1	2%
Non-evaluable, n (%)	1	2%
Median progression-free survival RECIST1.1, months (95% CI)	1.9 (1.8-2.0)	
Median progression-free survival iRECIST, months (95% CI)	1.9 (1.8-3.2)	
Duration of response, months (95% CI)	9.0 (4.7 – NR)	

Table S6: Characteristics of irradiation cohort. The irradiated lesions per patient, volume and dose of the irradiation, radiation technique and response of both the irradiated lesions and the other lesions (abscopal effect) after irradiation and the best response on nivolumab are depicted. In the rightmost column, the absolute change in stromal tumor-infiltrating lymphocytes (TIL) after irradiation is depicted for a biopsy from a non-irradiated lesion. Gy: Grey; CTV: clinical target volume; PTV: planning target volume; VMAT: volumetric modulated arc therapy; 3D-CRT: conventional radiotherapy with anterior-posterior fields or tangential fields; 3D-IMRT: intensity modulated radiotherapy.

Pa- tient ID	Irra- diated lesion	Dose	Vo- lume (mL)	Vo- lume type	Radia- tion tech- nique	Respon- s irra- diated lesion after irra- dia- tion	Respon- s non- irradiat- le- sions after irra- dia- tion	Best re- sponse irra- diated lesion after nivo- lumab	Best over- all re- sponse non- irra- diated le- sions after nivo- lumab	Absolute change in TIL after irradia- tion in biopsied lesion (non-ir- radiated)
5	Lymph nodes left medi- astinum	3 x 8 Gy	32.643	PTV	VMAT	Non- CR/non- PD	SD	Non- CR/non- PD	PR	-12.5%
7	Breast right	3 x 8 Gy	401.136	PTV	3D- CRT	Non- CR/non- PD	SD	Non- CR/non- PD	PD	+2%
9	Lymph nodes neck left	3 x 8 Gy	43.317	PTV	3D- IMRT	Non- CR/non- PD	Non- CR/non- PD	NA	PD (clini- cal)	-4.5%
17	Lymph nodes axilla left	3 x 8 Gy	45.2	PTV	VMAT	Non- CR/non- PD	SD	Non- CR/non- PD	PD	NA
28	Hip right	3 x 8 Gy	235.6	CTV	3D- CRT	Non- CR/non- PD	PD	NA	PD (clini- cal)	+4%
37	Lymph nodes supr- aclav- icular right	3 x 8 Gy	12.2	CTV	3D- CRT	Non- CR/non- PD	SD	Non- CR/non- PD	SD > 24 weeks +7%	
38	Right breast	3 x 8 Gy	635.86	CTV	3D- CRT	PD	PD	PR	PD	-2%
47	Axilla right	3 x 8 Gy	56.997	PTV	3D- IMRT	Non- CR/non- PD	SD	NA	PD (clini- cal)	-32.5%

53	Lymph nodes neck right	3 x 8 Gy	35.64	PTV	VMAT	Non- CR/non- PD	PD	NA (not im- aged)	PD	-2%
60	Liver seg- ment 6/7	3 x 8 Gy	129.64	PTV	VMAT	PD	PD	PD	PD	+2.5%
67	Lung middle lobe	3 x 8 Gy	38.628	PTV	VMAT	PD	PD	NA	PD (clini- cal)	NA
68	Sternum	3 x 8 Gy	64.38	PTV	VMAT	Non- CR/non- PD	SD	Non- CR/non- PD	PD	+4%

Part II

The effects of T cells on the tumor microenvironment

LACK OF DETECTABLE NEOANTIGEN DEPLETION IN TREATMENT-NAÏVE CANCERS

Maarten Slagter, Lorenzo F. Fanchi, Marit M. van Buuren, Arno Velds, Gergana Bounova, Lodewyk F.A. Wessels, and Ton N. Schumacher

bioRxiv: 10.1101/2023.06.21.544805

Abstract

While neoantigen depletion, a form of immunoediting due to Darwinian pressure exerted by the T cell based immune system during tumor evolution, has been clearly described in murine models, its prevalence in treatment-naïve, developing human tumors remains controversial. We developed two novel methodologies to test for depletion of predicted neoantigens in patient cohorts, which both compare patients in terms of their expected number of neoantigens per mutational event. Application of these strategies to TCGA patient cohorts showed that neither basic nor more extensive versions of the methodologies, controlling for confounding factors such as genomic loss of the HLA locus, provided statistically significant evidence for neoantigen depletion. In the subset of analyses that did show a trend towards neoantigen depletion, statistical significance was not reached and depletion was not consistently observed across HLA alleles. Our results challenge the notion that neoantigen depletion is detectable in cohorts of unmatched patient samples using HLA binding prediction-based methodology.

Introduction

Immune evasion, the avoidance of immune detection and eradication, is a hallmark of cancer^{1,2} and can occur through various mechanisms, including deletion of components of the antigen presentation machinery, insensitivity to pro-apoptotic or growth-arresting molecules, such as granzymes and IFN- γ ³, or expression of T-cell checkpoint ligands such as PD-L1. In addition, the Darwinian selective pressure exerted by T cells has been proposed to lead to outgrowth of tumor cells that lack T cell-recognized (neo-)antigens. In line with this, loss of a neoantigen that was recognized by T cells has been observed in a murine sarcoma model⁴. In addition, loss and reduced expression of mutant genes encoding T-cell recognized neoantigens has been observed in two patient case studies⁵, and the reduction in tumor mutational burden observed in clinical responders to PD-1 blockade has been proposed to lower the number of T cell-recognized neoantigens⁶. While collectively these data form relatively strong evidence that neoantigens can be lost upon (therapeutically enhanced) T-cell pressure, it is less clear whether such genomic neoantigen depletion of neoantigens widely occurs in treatment-naïve tumors, and – importantly - would be detectable at the genomic level. An argument against the idea that neoantigen depletion could be readily detected at the genomic level is formed by the observation that only a minor fraction of predicted neoantigens appears to naturally induce T-cell responses in patients⁷⁻¹⁰. Importantly, only this small subset of predicted neoantigens can be expected to be under Darwinian pressure, and this will affect the sensitivity of any methodology that examines the occurrence of predicted neoantigens regardless of the fact whether or not T cell reactivity was present against these predicted antigens.

Several prior studies have assessed neoantigen depletion in large sets of cancer genomes with unmatched samples (i.e., with a single tumor sample per patient) provided by the Cancer Genome Atlas (TCGA). A first of these studies provided evidence for the selective loss of mutations predicted to encode MHC-class I neoantigens for colorectal and clear cell kidney cancer¹¹, but at the same time observed a counter-intuitive neoantigen enrichment in EBV⁻ stomach adenocarcinoma. Likewise, application of a model of peptide immunogenicity found recurrent mutations to appear more readily in TCGA patients that are less capable of presenting the resulting new peptide sequences by HLA¹². In line with these results, an earlier TCGA pan-cancer analysis found recurrent oncogenic mutations to be relatively poorly HLA-presentable by the patients that carried them¹³, a result that has however since been shown to be fully driven by confounding factors^{14,15}. The same group also reported elevated neoantigen levels in tumors that harbor mutations in the antigen presentation pathway, but this result was not corrected for, potentially confounding, background mutation rates¹⁶. Using a mathematical model of tumor evolution, Lakatos et al. predicted the variant allele frequency (VAF)-spectrum of a tumor's somatic mutations under various degrees of immune pressure and found TCGA tumors to appear similar to simulations under immune pressure¹⁷. In contrast to the aforementioned studies, a comparison of the ratio of non-synonymous to synonymous mutation count ('dN/dS') between areas of the human genome that do or do not encode predicted HLA presented peptides found no evidence of negative selection when correcting for sequence con-

tent between these two classes of genomic regions¹⁸. Notably, this sequence content was shown to affect the ratio with which mutagenic processes yield either non-synonymous or synonymous mutations and so confounded dN/dS estimates¹⁸. A more recent study using dN/dS methodology did report depletion, especially of clonal mutations, in highly immune infiltrated tumor types and in patients that did not show other means of immune evasion, but did however not correct for sequence content¹⁹. This latter study also showed substantial degrees of nonsensical neoantigen enrichment, especially in lowly immune infiltrated tumor types.

Given the conflicting results in these prior analyses, there is a need for novel methodology to assess neoantigen depletion in treatment-naïve tumors. Here, we present two new, interrelated, methods for the detection of average neoantigen depletion in patient cohorts. Using these methods, we do not observe substantial evidence for neoantigen depletion in TCGA tumors, despite incorporation of many potential confounding co-variables into our modeling to increase sensitivity. We emphasize that the lack of signal that we describe does not rule out the occurrence of neoantigen depletion for a minority of studied patients and/or a minority of (T cell-recognized) neoantigens. However, our observations do challenge the notion that neoantigen depletion signals based on HLA affinity predictions are detectable in large-scale unmatched cancer sequencing data sets.

Results

Design and evaluation of an HLA-I antigen prediction pipeline

To study Darwinian selection against neoantigenic non-synonymous mutations, we first developed an (neo-)epitope prediction pipeline and optimized it with respect to prediction precision and sensitivity. The pipeline annotates candidate nonameric peptides with the output of four tools that jointly model the major requirements for (neo-)antigen presentation: RNA expression of the mutant DNA sequence, predicted proteasomal processing, predicted HLA-binding, and self-similarity of encoded peptides (Figure S4.1A). To tune the parameterization of this pipeline, we first identified a set of peptides within the 3,094 nonameric peptides present in the HIV genome for which T-cell recognition in the context of the HLA-A*02:01 allele, the most common HLA class I allele in US and European populations, had been demonstrated unambiguously. Specifically, by querying the HIV Molecular Immunology Database²⁰, we identified peptides that met the following criteria: (i) HLA subtyping information had demonstrated restriction by the HLA-A*02:01 subtype; (ii) T-cell responses had been observed in at least 3 patients; and (iii) such T-cell responses had been observed in material from HIV-infected patients, rather than induced in vitro or in animal models. This resulted in a set of 32 epitopes in the HIV genome for which presentation by HLA-A*02:01 and recognition by the human T-cell repertoire had unambiguously been established (Table S2). Subsequently, we used this epitope set to compare the sensitivity and precision of epitope prediction strategies that either solely relied on predicted HLA binding affinity²¹, or that also integrated filters that predict protea-

somal processing²² and similarity of a candidate epitope to self-peptides predicted from the human proteome²³, methods). As compared to epitope predictions solely based on predicted HLA affinity, we found both incorporation of proteasomal processing and exclusion of peptides with similarity to human sequences to improve prediction precision (Figure S4.1B), in line with other work²⁴. Selection of an affinity threshold of 255 nM yielded a good balance between precision and sensitivity (Figure S4.1B-D) of this prediction pipeline. We note that experimental data on T-cell recognized HLA-A*02:01 peptides in the HIV genome may still be incomplete, rendering these performance estimates lower bounds on true prediction performance.

To complement the HIV-pipeline validation, we did a similar validation on peptides derived from the IEDB for 10 viruses. The coverage of T-cell recognized peptides in the IEDB database was likely substantially lower, as in this database we retrieved only 9 out of the 33 T-cell targeted HIV-peptides that were identified in the Los Alamos database, and this probably underlies the observation that much stricter pMHC affinity-filtering yielded optimal precision for the IEDB derived epitope set (Figure S4.1E,F).

The two patient-group methodology

To next determine whether neoantigen depletion detectably shapes the mutational landscape of human cancer during tumorigenesis, we first compared the number of predicted HLA-A*02:01 neoantigens per mutational event (the neoantigen yield rate, r) for tumors of HLA-A*02:01 positive patients (test set) and tumors from patients that lack the HLA-A*02:01 allele and also other HLA class I alleles with similar peptide binding profiles (reference set, see below for the filtering of HLA alleles with similar peptide binding profiles). As the latter patient group cannot present these predicted HLA-A*02:01 restricted neoantigens, this group provides a reference that can be used to calculate to what extent T-cell pressure has shaped the repertoire of neoantigens as predicted on tumor mutation data (Figure 4.1A). In addition, by focusing on neoantigen predictions for a single class I HLA-allele, rather than the diverse set of alleles carried by any individual patient, an equal and, in case of HLA-A*02:01 high prediction accuracy is guaranteed across patients.

To be able to assign patients to either the test or reference groups, we first assessed the similarity in peptide binding properties between HLA-A*02:01 and all other HLA class I alleles encountered within our patient set, by computing for each observed HLA-allele the fraction of the 62,833 predicted HLA-A*02:01 neoantigens observed in this cohort that it was also predicted to present (corroboration index, Figure S4.1A). HLA class I alleles that were predicted to present more than 20% of the set of HLA-A*02:01 peptides were classified as ‘HLA-A*02:01like’ (37 of 224 HLA class I alleles). Subsequently, patient samples expressing at least one such allele were removed from the analysis, resulting in an HLA-A*02:01 positive test group of 2,345 patients and a reference group of 2,628 patients that lacked HLA-A*02:01 and also any HLA class I allele with substantially overlapping binding properties. Residual binding repertoire overlap between the groups was small, with reference group

patients being able to present 16.7% (median) of their predicted HLA-A*02:01 presented peptides using any of the up to 6 HLA class I alleles that they expressed (Figure 4.11B), i.e., 6-fold lower than in the HLA-A*02:01 positive test group.

To test for preferential loss of mutations that yield predicted HLA-A*02:01 presented neoantigens, we compared r between test and reference patient groups per tumor type. This analysis revealed a significantly lower r in the test set for ovarian carcinoma, while a higher r in the test set, i.e., a presumed neoantigen enrichment, was observed for glioblastoma. However, neither of these results remained significant after multiple testing correction (Figure 4.1C, left), indicating a lack of detectable neoantigen depletion above the noise levels in the data. Absence of significant depletion was also observed when aggregating all mutations for each tumor type into one HLA-A*02:01 positive and one HLA-A*02:01 negative ‘meta-patient’, an approach that is expected to increase analysis robustness in case of low mutation numbers (Figure 4.1C, right).

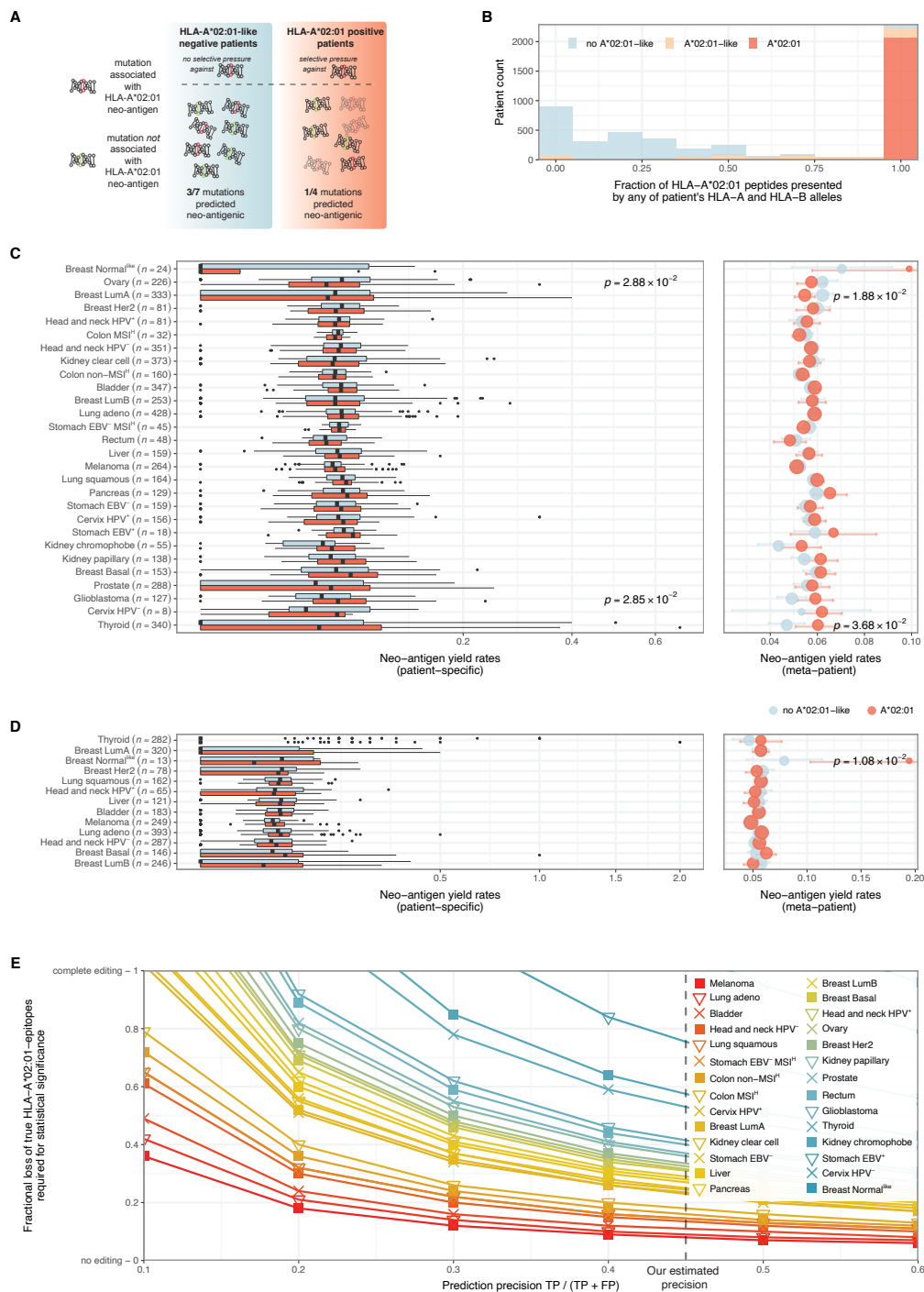


Figure 4.1: No detectable genomic loss of predicted HLA-A*02:01 neoantigens using a discrete two-group detection strategy.

A Schematic overview of analysis strategy. Mutational events that yield HLA-A*02:01-restricted neoantigens can undergo negative selective pressure in HLA-A*02:01-positive patients (the test group), but not in patients that lack this allele as well as other HLA class I alleles with similar peptide binding profiles (“HLA-A*02:01-like negative reference patients”). As a consequence, genomic loss of antigenic HLA-A*02:01 mutations would be reflected in a lower HLA-A*02:01 neoantigen yield rate in HLA-A*02:01-positive patients as compared to HLA-A*02:01-like negative patients (in the indicated example, 3 out of 7 vs. 1 out of 4 non-synonymous mutations yield predicted HLA-A*02:01 neoantigens).

B Distribution of the fraction of HLA-A*02:01 binding peptides presented by any of a patient’s HLA class I A and B alleles, according to HLA-A*02:01-allele and HLA-A*02:01-like allele status. Patients in the HLA-A*02:01-like negative reference group present 16.7% (median) of predicted HLA-A*02:01 binding peptides by any of their HLA class I A and B alleles.

C Neoantigen yield rates of SNVs in HLA-A*02:01-positive and HLA-A*02:01-like negative tumor samples. Left: distributions of patient specific neoantigen yield rates. Right: ‘meta-patient’ yield rates, in which mutations, and hence neoantigens, are grouped per tumor type. Unadjusted p-values are shown for tumor types where they are smaller than or equal to 0.05. None of the comparisons remained significant after correcting for multiple testing.

D As in C., but excluding both subclonal mutations and mutations for which loss may potentially confer a cell intrinsic fitness cost (driver mutations and mutations in essential genes that display LOH). Unadjusted p-values are shown for tumor types where they are smaller than or equal to 0.05. None of these remained significant after correcting for multiple testing.

E Power analysis of neoantigen depletion detection strategy. The effect of prediction precision on the fractional loss of predicted neoantigenic mutations required to achieve statistical significance is depicted on a tumor type-specific basis. Dashed vertical line depicts precision of the here employed epitope prediction pipeline, indicating that genomic editing of a minimum of 9% and 18% of true HLA-A*02:01 neoantigens would have been detected in melanoma and colon MSI^H, respectively, assuming no multiple testing correction of p-values.

A potential limitation of this analysis strategy could be formed by the presentation of a fraction of HLA-A*02:01presentable peptides by the aggregate of all the other HLA-alleles in the reference set patients, such that a degree of ‘background’ immune pressure can be expected to occur against HLA-A*02:01peptides. To test robustness with respect to the stringency used to create the non-HLA-A*02:01-like reference patient set, we performed meta-patient tests using an increasingly strict HLA-similarity threshold (i.e., allowing a progressively lower fraction of HLA-A*02:01presented peptides to be presented by the non-HLA-A*02:01-like reference patient set). Using this approach, no systematic increase in statistical significance was observed (Figure S4.1B), suggesting that the (low-level) overlap in peptide HLA-binding profiles was unlikely to confound this analysis.

As the probability of epitope presentation depends on the RNA expression level of the associated gene²⁵⁻²⁸, an increase in RNA expression thresholds for neoantigen predictions may be expected to increase precision (i.e., increase the fraction of truly presented peptides amongst predicted peptides), and could thereby potentially reveal a weak neoantigen depletion signal. Similarly, loss of neoantigens could be postulated to be more apparent among high affinity HLA ligands that are more likely to yield strong T cell targets. To test both possibilities, we titrated both the RNA expression and HLA affinity thresholds of the neoantigen prediction pipeline and re-evaluated HLA-A*02:01 *r* in the test and reference patient sets. Small differences between these two sets were observed when performing pipeline stringency titrations and we did observe significantly lowered *r* for multiple analyses

(i.e., rectum, breast LumB, melanoma, head and neck HPV⁺, Figure S4.3C). Importantly however, a similar number of tumor types with a heightened r was once again observed (Figure S4.3C). We conclude that absence of detectable neoantigen depletion is robust to different configurations of the neoantigen prediction pipeline.

To screen for other factors that could have obscured a weak neoantigen depletion signal, we performed additional analyses in which we excluded mutations for which negative selection could have been counter-acted by positive selection. Specifically, T cell-mediated depletion of oncogenic mutations in driver genes may be expected to be counteracted by the positive effect of these mutations on cellular fitness²⁹. Similarly, essential genes that encode neoantigens and of which the wild type copy has been lost (so called essential passengers²⁹) cannot be lost without loss of cell viability^{29,30}. Finally, we performed analyses that excluded subclonal mutations, as these could be postulated to have emerged too recently in order for the immune system to have affected their presence. When excluding the aforementioned mutation classes, we again discerned no statistically significant neoantigen depletion, neither when analyzing patient-specific nor meta-patient neoantigen yield rates (Figure 4.1D, left and right, respectively, fewer tumor types due to lower availability of required data for clonality calling, Methods).

In prior work, the HLA class I-restricted presentation of neoantigens has been reported to shape the repertoire of oncogenic mutations, with individual driver mutations being reported to occur more frequently in patients with HLA repertoires that are less likely to present the resulting mutant peptides¹³. We attempted to validate this observation using our epitope prediction and analysis strategy. Restricting our analysis to the recurrent driver single nucleotide variants (SNVs)¹³, we also did not observe significant neoantigen depletion (Figure S4.3D). Finally, escape from immune pressure can occur through a variety of genetic alterations, including mutations in components of the antigen presentation machinery, and the loss of T-cell-recognized neoantigens can reasonably be expected to no longer provide a fitness advantage in tumors that harbor such alterations. To restrict our analysis to tumor samples for which no evidence of other known escape mechanisms was present, we excluded all tumors with one or more non-silent mutations in any of 515 genes implicated in resistance to T-cell killing through CRISPRi screening³¹. Application of this tumor sample filtering to all of the analyses reported above resulted in larger differences for at least one tumor type per analysis. However, this did not result in an increased number of tumor types for which significant neoantigen depletion was observed after correcting for multiple testing (Figure S4.3D). Our recurrent finding of absence of significant neoantigen depletion in treatment-naïve tumors contrasts with an earlier assessment of neoantigen depletion on TCGA data¹¹. Adapting this prior analysis strategy to our HLA-A*02:01 centered neoantigen predictions (Methods) again did not reveal any signs of genomic epitope loss (Figure S4.3E), even when restricting this analysis to clonal mutations (Figure S4.3F), or when exhaustively testing other possible variants of this analysis (Figure S4.3D).

Even when using optimized epitope prediction pipelines, neoantigen predictions still contain a sizeable fraction of false positive peptides. As only true positive peptides (i.e., predicted peptides that are

true HLA ligands) can potentially be seen by T cells, and hence can be subject to Darwinian selection, prediction precision defines the lower limit for the degree of neoantigen depletion that would be detectable. We determined what degree of neoantigen depletion would have been detectable using the two-group methodology, by computing the required effect size (i.e., fractional loss of true epitopes in HLA-A*02:01 positive tumors) to reach statistical significance given the observed noise in the data. At our estimated neoantigen prediction precision of 0.45 (Figure S4.1B,C), unadjusted p -values of 0.05 would have been reached upon loss of 8-18% of truly presented epitopes in melanoma, lung and colon cancers (Figure 4.1E), indicating that the true neoantigen depletion signal must have been smaller than that for it to have not been detected.

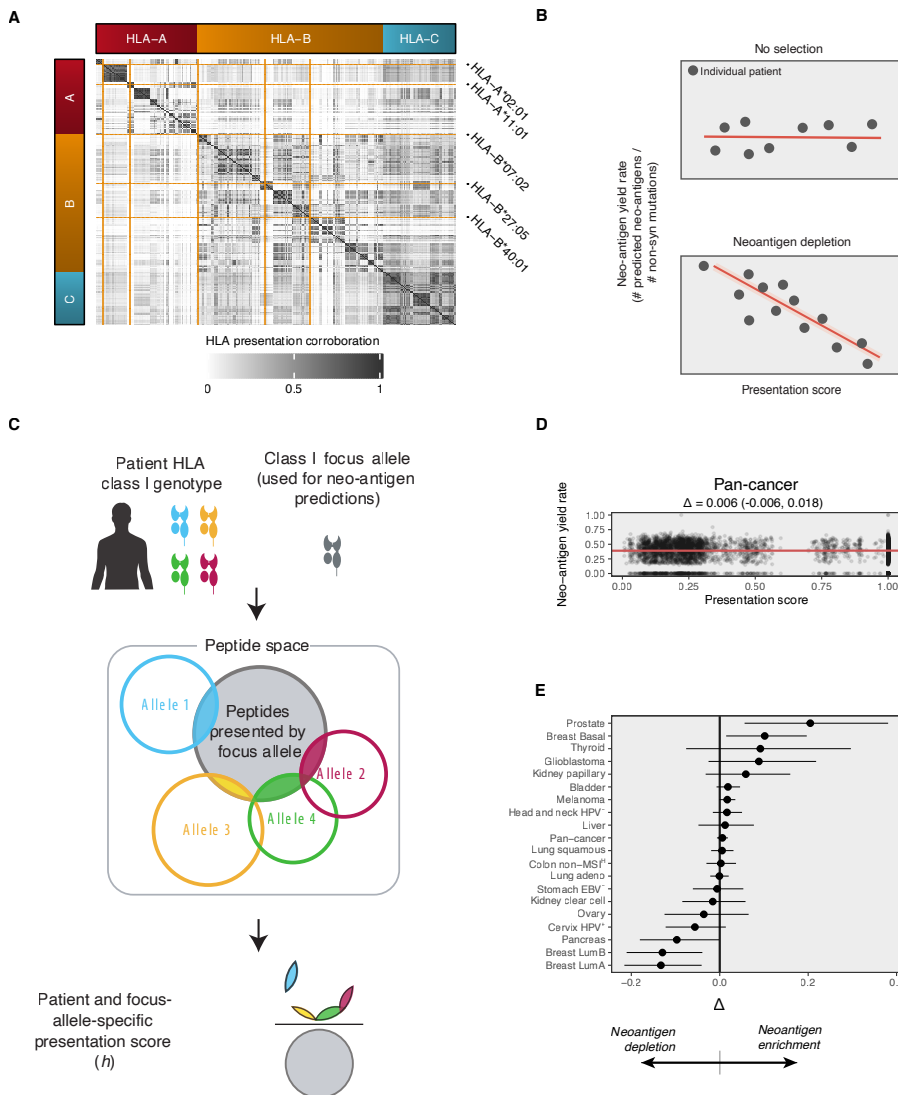


Figure 4.2: continuous strategy to detect neoantigen depletion

A Overlap in predicted peptide presentability across all detected HLA class I alleles (A, B and C) across the TCGA patient samples included in the analyses. Matrix reflects the fraction of peptides presented by alleles in the rows (i.e., NetMHCpan 3.0 rank percentile ≤ 1.9), that are also presentable by the alleles in the columns (same presentation criteria). Alleles in both rows and columns are alpha-numerically ordered. The five highlighted HLA-A and -B alleles were selected based on NetMHCpan3.0 prediction accuracy, HLA diversity and to minimize overlap in predicted binding capacity within the group.

B Schematic of the continuous (as opposed to two patient-group) methodology to evaluate neoantigen depletion. To account for overlap in (predicted) HLA binding, a continuous score was used that captures a sample's capability of HLA-presenting peptides that are associated with an HLA allele of interest (the 'focus allele'). This score is then regressed against the neoantigen yield rate for this focus allele. A negative slope in this regression would indicate that fewer neoantigen-encoding mutations are detectable in samples that have a high capacity to present these.

C Schematic of the HLA presentation score (h), reflecting the fraction of peptides presentable by the focus allele, that is also presentable by at least one of the HLA alleles carried by a patient.

D Pan-cancer result with a basic parameterization of the analysis, as in Figure 1C.

E Forest plot of the observed ϕ values and associated 80% confidence intervals.

A continuous version of the neoantigen depletion analysis

Given the overlap in binding profiles of different HLA class I alleles (Figure 4.2A), and the resulting continuous rather than bi-modal distribution of HLA-presentation overlap with HLA-A*02:01 across samples (or any other HLA-allele, hereafter called the focus allele), a statistically more powerful approach could be to test for a (negative) association between focus allele-presentation capability and neoantigen yield rate (r) across samples. Modelling focus-allele presentation capability with a quantity we call the HLA presentation score (h , methods), a detectable signal of neoantigen depletion (i.e., depletion of neoantigenic, non-synonymous mutations), should on average lead to a lower r in samples with a high h . That is, a linear regression between h and r should yield a negative slope (Figure 4.2B, bottom). In contrast, a slope of zero would indicate HLA presentability not to be associated with depletion of mutations carrying predicted neoantigens (Figure 4.2B, top). We modeled h as the fraction of unique, theoretically presentable focus allele peptides that are also presentable by one or more of the patient's class I alleles (Figure S4.2C). In this way, h will be 1 for patients that do carry the focus allele while those that do not will have values ranging from 0 to 1.

To first evaluate whether h models peptide presentation capacity, we analyzed mass spectrometry data of HLA-eluted peptides³². Specifically, having inferred the HLA class I repertoire of each patient in this data set, we computed h for each patient and potential focus allele combination. Next, for all 9-mer peptides of the expressed human proteome we predicted whether mass spec detection would have been expected solely based on peptide affinity predictions. In case h models peptide presentation correctly, the proportion of predicted peptides for a given focus allele to be detected by mass-spec should be correlated to h . Confirming our expectation, we observed positive correlations between h and the number of detected over expected peptides for 23 of 27 analyzed samples (Figure S4.4). Having computed h across all five focus alleles and all samples, we observed a near-zero

correlation between focus alleles (Figure S4.5A,B). This indicates that analyses using different focus alleles form largely independent and complementary tests within a fixed set of tumor samples, and can hence be seen as independent validations.

We next assessed the relationship between h and SNV r across all evaluated patients (i.e., pan-cancer) using HLA-A*02:01 as the focus allele, observing a virtually flat, non-statistically significant, slope. To put this slope into perspective, we defined Δ (delta) as the relative difference in r between patients at $h = 0$ (no ability to present focus-allele peptides) and $h = 1$ (full ability to HLA-present focus-allele peptides, Methods), and observed a Δ of 0.006 (80% CI: [-0.006, 0.018], i.e., non-significant enrichment rather than depletion of predicted neoantigenic mutations, Figure 4.2D). Testing of individual tumor types in this manner showed apparent neoantigen depletion in certain tumor types ($\Delta < 0$), but just as many other types showed a similarly strong apparent enrichment for neoantigens ($\Delta > 0$), likely reflecting noise in the data (Figure 4.2E).

Analogous to the two-group based neoantigen depletion methodology described above, we next systematically varied all possible settings of the neoantigen prediction pipeline and the continuous neoantigen depletion analysis strategy in order to test the robustness of these results. As the optimal neoantigen prediction pipeline configuration remains uncertain, despite our efforts to determine it (Figure S4.1B-F), we evaluated all outcomes while varying three settings of the neoantigen prediction pipeline: i) the *HLA affinity rank percentile threshold* that denotes the predicted HLA affinity candidate neoantigens had to reach for predicted HLA-presentation (4 levels, ranging from lenient to highly stringent), ii) *RNA expression*, either evaluated at the gene level or at the variant level (the latter to be sensitive to potential epigenetic silencing of neoantigenic mutations) and iii) the *similarity-to-self* filter that aims to model thymic selection of T-cell reactivity.

We also varied 6 settings that determine how the neoantigen depletion test is performed, independent from the neoantigen prediction pipeline configuration: i) *Variant selection* determines the set of somatic variants based on which r was evaluated and included the following classes: all (SNV) genomic variants; frameshifting indels as these may form richer sources of neoantigens that are typically less self-similar^{33,34} and could thereby be postulated to experience stronger negative selection; only clonal SNVs mutations; SNVs with driver and essential passenger mutations removed²⁹; only highly-recurrent driver SNVs, as defined in a prior work¹³. ii) *Focus-allele* reflects the HLA-allele for which both h and r was evaluated (the two main variables of the regression). The focus allele was varied between five HLA class I-alleles that we picked to cover a broad range of HLA super-types and that showed a relatively high prediction accuracy using NetMHCpan3.0²¹. iii) *LOH in HLA* determines whether allelic loss of HLA class I³⁵ was reflected in the presentation score (h), i.e., whether HLA alleles that were reliably found to have been genomically deleted were or were not excluded when computing h . We discerned between a high-confidence variant of this variable, in which only patients for which all alleles could be reliably assessed were included in the final regression analysis (“strict LOH HLA”, Figure S4.6, Methods), and a more lenient version in which any allele reliably found to be lost was excluded from the HLA presentation score alleles, independent

of the assessment quality for other alleles in the same sample (Methods). iv) *C-allele in presentation score* reflects whether both HLA-A, -B, and -C alleles, or only the HLA-A and -B alleles, were included when computing *h*. As most known neoantigens are HLA-A or HLA-B restricted³⁶ and HLA-C might be expressed at a lower level³⁷, exclusion of HLA-C could lead to more accurate estimates. v) *T cell-resistance*, entails the removal of patient samples from the analysis that carry one or multiple mutations, other than genomic loss of HLA class I loci, in immune evasion genes as identified using CRISPRi-screening³¹, in order to exclude samples in which immune pressure may be reduced by an independent genetic event. Here, patient samples were excluded using three levels of stringency, reflecting false discovery rate thresholds for genes in the CRISPR-screen of 0.0001 (stringent on genes and hence lenient on sample inclusion), 0.1 (moderate on sample inclusion) and 0.01 (stringent on sample inclusion). vi) *Cytolytic score* determines whether sample inclusion is restricted to samples with a high T cell infiltration, as neoantigen depletion might be more apparent in immune infiltrated tumors¹⁷. here we restricted the tested samples to those high ($\geq 75^{\text{th}}$ quantile within the tumor type) in cytolytic score¹¹ (a transcriptomic proxy for T/NK-cell activity) and those lower in cytolytic score ($< 75^{\text{th}}$ quantile) to contrast the former results with.

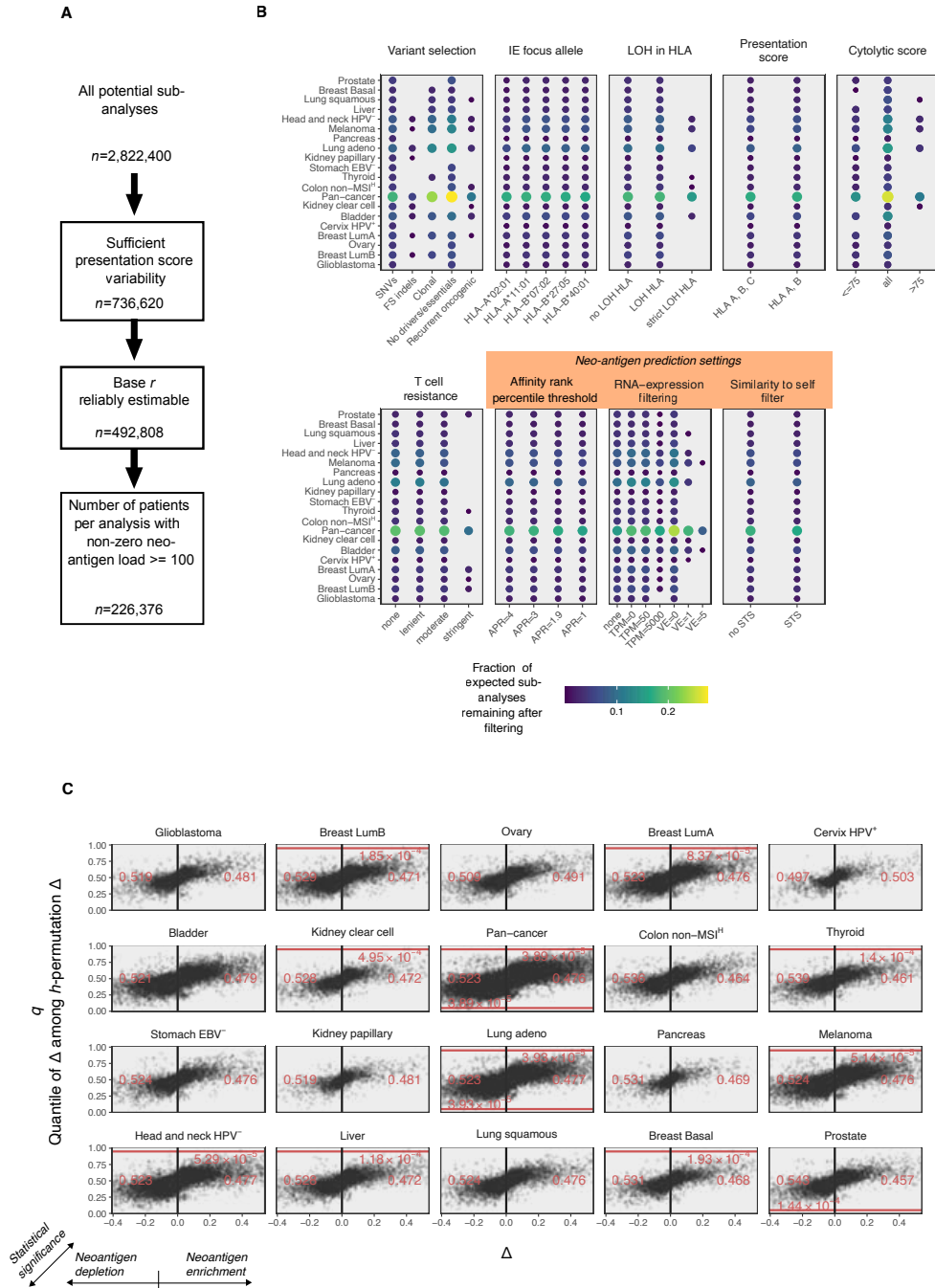


Figure 4.3: Extended search for neoantigen depletion using the continuous detection methodology
A Filtering of sub-analyses to enrich for sub-analyses with acceptable inferential quality. Blocks denote filtering steps, reported numbers are the number of remaining sub-analyses after filtering.
B Sub-analysis composition differs between tumor types because of technical (e.g., the number of patient samples, the quality of DNA/RNA-sequencing data) and biological (mutational load) reasons. Chart shows the frac-

tion of times each factor level (horizontal axis) is observed among the filtered sub-analyses. Levels that lead to strong reductions in predicted neoantigen counts or severely limit sample numbers are only feasible for high mutational load/ more highly represented tumor types.

C Volcano-like plot of all filtered sub-analyses, each dot represents a single sub-analysis. Due to compositional differences (see panel B), these plots cannot be used to directly compare tumor types to each other. In case of presence of eccentric sub-analyses ($q < 0.05$ or $q > 0.95$), red lines highlight these thresholds and the fraction of sub-analyses surpassing these are indicated in red.

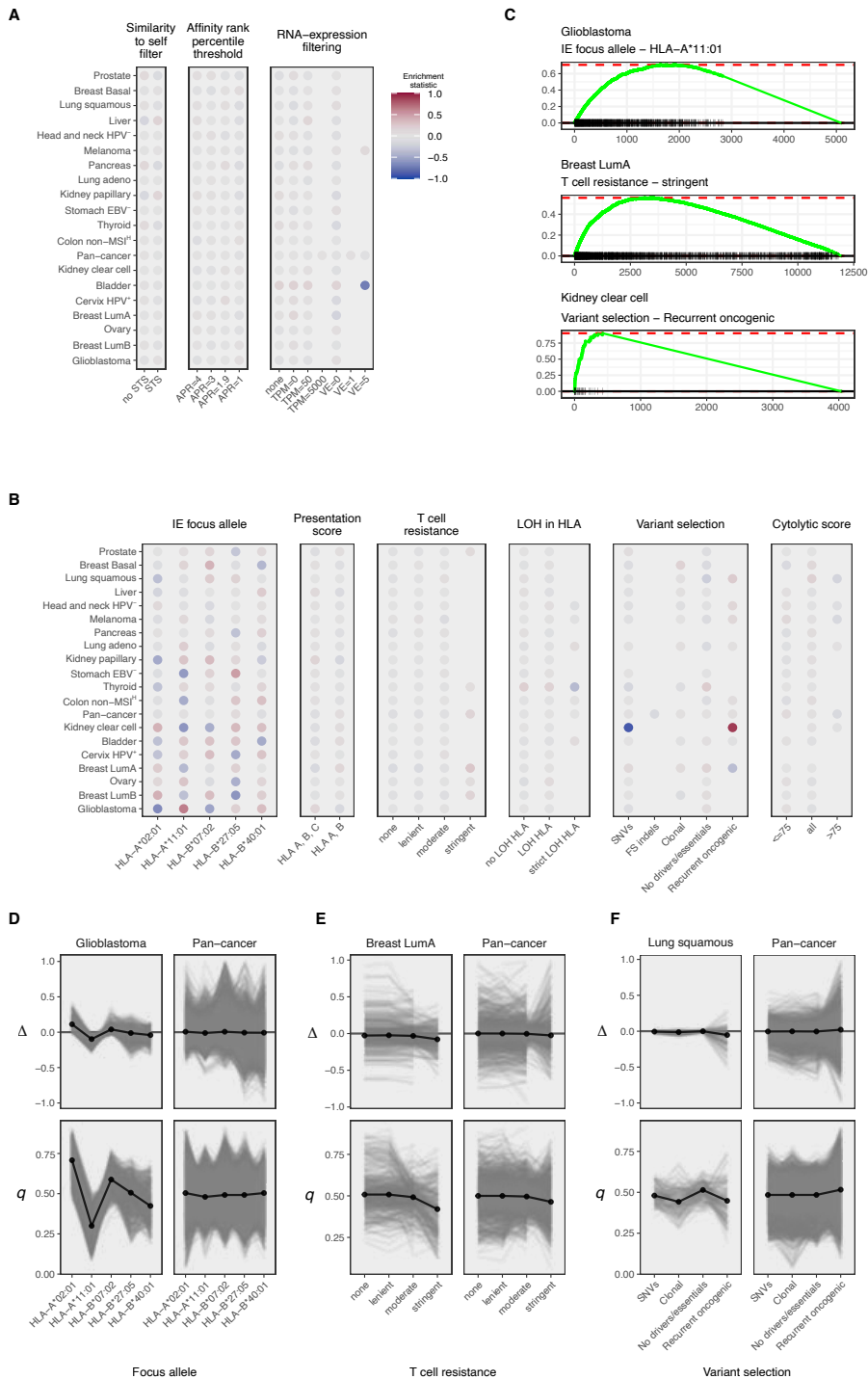


Figure 4.4: Extended analysis of continuous methodology to assess neoantigen depletion

A Enrichment scores for sub-analysis settings associated with neoantigen prediction methodology, after rank-

ing the sub-analyses by q . Note that none of the settings is strongly associated with q . ID-variable levels appearing in less than 10 sub-analyses were excluded from analysis.

B As in A. but for settings associated with the manner in which neoantigen depletion analysis is performed. As the choice for a focus allele is arbitrary, it is not expected to affect the resulting statistics in case of a detectable and true signal, but it is shown here to be the dominant source of variation. Color scale as in B.

C Highlighted individual enrichment-analyses from B, selected for their relatively strong enrichment scores.

D Randomly selected groups of identically parameterized sub-analyses, differing only in focus allele, to directly visualize the average effect of variation in focus allele. Glioblastoma and pan-cancer analyses are highlighted to illustrate the relatively strong and weak enrichment scores, respectively, for these tumor types in panel C.

E As in D., but varying the subselection of potentially T-cell resistant samples on the horizontal axis. ‘Stringent’ denotes the most rigorous filtering of tumor samples based on the presence of potential immune evasion mechanisms, selecting tumor samples with minimal numbers of mutations that have been associated with immune evasion.

F As in D., but varying the somatic variant set on which r is evaluated. Frameshift indels were excluded as they could only be reliably assessed on the pan-cancer level (in which they weren’t associated with q or Δ).

Evaluating all combinations of the settings listed above, we frequently encountered patient subsets that became prohibitively small for regression analysis (i.e., uncertain or no regression coefficients) and neoantigen yield rates that became prohibitively low (resulting in a majority of patients with 0 predicted neoantigens). Restricting ourselves to combinations (so called ‘sub-analyses’) for which i) at least 25 samples had h in the range $[0, .25]$ and in the range $[0.75, 1]$, ii) the baseline r , i.e., r for samples with $h = 0$, could be reliably estimated by the model (methods), and iii) at least 100 samples had non-zero neoantigen loads, left 226,376 of 2,822,400 theoretically possible sub-analyses (8.0%, Figure 4.3A). As expected, tumor types with a large number of samples allowed more restrictive sample filtering and those with high TMB allowed more stringent neoantigen filtering. As such, individual levels of settings were present with variable frequency between tumor types, such that tumor types could not directly be compared from the set of all filtered sub-analyses (Figure 4.3B). Assessing Δ across the filtered sub-analyses, we did discern tumor types for which the distribution appeared (strongly) skewed to the negative side (Figure 4.3C, consistent with neoantigen depletion). However, this distribution simultaneously also appeared to be positively skewed for other tumor types. To assess the statistical significance of these and other sub-analyses, we employed permutation tests. Specifically, for each sub-analysis, we permuted h -scores across individual samples in the source data 250 times, evaluated Δ for each of these permutations, and assessed the fraction of permutation Δ s that was smaller than the original Δ , arriving at a quantity we call q (for quantile). As q virtually never reached below .05 ($n = 4$, 0.002%) or above .95 ($n = 11$, 0.005%, Figure 4.3C, vertical axis), shifts towards either negative or positive Δ (i.e., neoantigen depletion and neoantigen enrichment) did not appear statistically significant.

To characterize these results and identify which individual settings most strongly enriched for neoantigen depletion, we ordered all sub-analyses by q and used preranked gene set enrichment analysis³⁸ to identify sub-analysis settings that were associated with extreme values in q in a univariate manner. As all of the settings are categorical, we compared individual levels (e.g., HLA-A*02:01 for the focus allele) against the combination of all other levels. In this, we recorded the normalized loca-

tion (range: $[-1, 1]$) where the rank sum statistic deviated most strongly from that expected under random ordering (i.e., the enrichment statistic³⁸). Using this strategy, we did not find evidence that more stringent (and precise) neoantigen prediction could reveal neoantigen depletion (Figure 4.4A). Surprisingly, repeating this analysis but sorted by Δ rather than q showed that stringent RNA expression filtering on the variant level on average led to a negative Δ for a majority of tumor types (Figure S4.7A). This was however likely caused by the combination of a heterogeneous distribution of h combined with low r (Supplemental Note 1), and hence artefactual.

When evaluating all settings of the neoantigen depletion testing procedure, we observed that the focus allele most strongly affected both q (Figure 4.4B-D) and Δ (Figure S4.7D). Varying the choice of the focus allele allows for semi-independent replication of a neoantigen depletion test, as there is no expected link between the arbitrary choice of an HLA class I allele as the focus allele and the occurrence of neoantigen depletion or enrichment. Importantly, the large observed variation between sub-analyses carried out in this manner and that are otherwise identically parameterized suggests that most of the signal observed can be ascribed to measurement noise. By the same token, we found rigorous exclusion of T cell-resistant tumors to shift q towards neoantigen depletion in a slight but consistent manner (5/5 evaluable tumor types, Figure 4.4B,C,E), but for only 1 out of 5 tumor types (Breast LumA, Figure S4.8A) this effect was somewhat consistent across the 5 evaluated focus alleles, again suggesting this effect to be spurious. Finally, evaluating r strictly on a set of 1018 recurrent oncogenic mutations that have previously been reported to occur preferentially in patients that express HLA class I alleles that are predicted to present these relatively poorly¹³, also might weakly enrich for a signal of neoantigen depletion (7/9 evaluable tumor types, Figure S4.4B,C,F), but this did not hold up at the pan-cancer level and - importantly - was highly inconsistent between focus alleles for all these 9 tumor types (Figure S4.8B), again indicating spurious associations.

Discussion

Using two different methodologies to estimate neoantigen depletion, we observed little if any detectable signal across unmatched tumor samples of treatment-naïve patients, also when controlling for a number of potentially confounding factors. This finding contrasts with part of the existing literature on the topic, in which evidence for neoantigen depletion in treatment naïve tumors was reported. In cases in which we tried to adapt existing methodologies to ours, we also did not detect neoantigen depletion or only observed minimal, non-statistically significant, trends.

It is important to emphasize that we do not see our data as evidence that neoantigen depletion does not occur. Specifically, given the support for immunosurveillance of nascent tumors³⁹ and the strong evidence favoring a role for T cell recognition of mutation-induced neoantigens in tumor control^{40,41}, we deem positive selection of tumor cell clones that never acquired or that lost T cell-recognized neoantigens plausible a priori. If neoantigen depletion does indeed occur, how can this be

reconciled with the observation that neoantigen depletion is not observable in the genomic analyses presented here? First, our data are consistent with the possibility that a small subset of neoantigens in pre-treatment tumors is lost due to immune pressure, below the level that surpasses the large prediction noise present in our analyses but also in the other approaches that have been explored. Specifically, experimental data suggest that only a small minority of predicted neoantigens (approximately 1%) induces detectable T cell responses^{8,42,43}. If only this small minority of T cell-recognized neoantigens is at risk of deletion, the maximal depletion signal that could be expected would be proportionally smaller, and would be difficult to detect in the background of predicted epitopes that never led to T cell pressure. On a related note, it is possible that neoantigen depletion does occur at the genomic level, but only or predominantly in later disease stages, or post-(immuno)therapy. Third, neoantigen depletion may be infrequent due to the presence of other, perhaps more potent or more easily accessible, mechanisms of immune evasion^{35,44}. Notably, across tumor types, the median fraction of patients harboring any form of genetic immune escape other than neoantigen loss in their primary tumors was reported to be 0.20 (for metastatic tumors: 0.27) and as high as 0.74 for kidney chromophobe cancer⁴⁴.

To increase the sensitivity of future analyses of genetic information on compendia of tumor samples, it will likely be critical to account for differences in immunogenicity between predicted neoantigens, rather than merely filtering mutations for those expected to be neoantigenic and implicitly assuming equal immunogenicity. At present, technologies to predict the development of T cell responses against a collection of predicted HLA-presented neoantigens are still limited in reliability, despite substantial efforts⁴⁵. Furthermore, if part of this process is stochastic, for instance governed by the occurrence of a specific TCR recombination, the development of accurate predictors of immunogenicity may be challenging. A more attractive approach may thus be to identify very large numbers of T cell recognized neoantigens in prospective clinical studies and to use such epitopes as a wet lab-validated starting point. High throughput T cell repertoire sequencing of T cells with tumor-reactivity signature⁴⁶, along with advances in experimental approaches to screen the reactivity of many T cell clonotypes in parallel, is expected to enable this effort⁴⁷⁻⁵¹.

Bibliography

1. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* *144*, 646–674. 10.1016/j.cell.2011.02.013.
2. Teng, M.W., Galon, J., Fridman, W.H., and Smyth, M.J. (2015). From mice to humans: developments in cancer immunoediting. *J Clin Invest* *125*, 3338–3346. 10.1172/JCI80004.
3. Zaretsky, J.M., Garcia-Diaz, A., Shin, D.S., Escuin-Ordinas, H., Hugo, W., Hu-Lieskovan, S., Torrejon, D.Y., Abril-Rodriguez, G., Sandoval, S., Barthly, L., et al. (2016). Mutations Associated with

- Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* 375, 819–829. 10.1056/NEJMoa1604958.
4. Matsushita, H., Vesely, M.D., Koboldt, D.C., Rickert, C.G., Uppaluri, R., Magrini, V.J., Arthur, C.D., White, J.M., Chen, Y.-S., Shea, L.K., et al. (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* 482, 400–404. 10.1038/nature10755.
 5. Verdegaal, E.M., de Miranda, N.F., Visser, M., Harryvan, T., van Buuren, M.M., Andersen, R.S., Hadrup, S.R., van der Minne, C.E., Schotte, R., Spits, H., et al. (2016). Neoantigen landscape dynamics during human melanoma-T cell interactions. *Nature* 536, 91–95. 10.1038/nature18945.
 6. Riaz, N., Havel, J.J., Makarov, V., Desrichard, A., Urba, W.J., Sims, J.S., Hodi, F.S., Martin-Algarra, S., Mandal, R., Sharfman, W.H., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* 171, 934–949 e15. 10.1016/j.cell.2017.09.028.
 7. Bentzen, A.K., Marquard, A.M., Lyngaa, R., Saini, S.K., Ramskov, S., Donia, M., Such, L., Furness, A.J., McGranahan, N., Rosenthal, R., et al. (2016). Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* 34, 1037–1045. 10.1038/nbt.3662.
 8. Karpanen, T., and Olweus, J. (2017). The Potential of Donor T-Cell Repertoires in Neoantigen-Targeted Cancer Immunotherapy. *Front Immunol* 8, 1718. 10.3389/fimmu.2017.01718.
 9. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128. 10.1126/science.aaa1348.
 10. van Rooij, N., van Buuren, M.M., Philips, D., Velds, A., Toebes, M., Heemskerk, B., van Dijk, L.J., Behjati, S., Hilkman, H., El Atmioui, D., et al. (2013). Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol* 31, e439–42. 10.1200/JCO.2012.47.7521.
 11. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. 10.1016/j.cell.2014.12.033.
 12. Schmidt, J., 1, 2, Harari, A., Bassani-Sternberg, M., and Gfeller, D. (2021). Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Rep. Med. in press*, 100194. 10.1016/j.xcrm.2021.100194.
 13. Marty, R., Kaabinejadian, S., Rossell, D., Slifker, M.J., van de Haar, J., Engin, H.B., de Prisco, N., Ideker, T., Hildebrand, W.H., Font-Burgada, J., et al. (2017). MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* 171, 1272–1283 e15. 10.1016/j.cell.2017.09.050.

-
14. Claeys, A., Luijts, T., Marchal, K., and Eynden, J.V. den (2021). Low immunogenicity of common cancer hot spot mutations resulting in false immunogenic selection signals. *PLOS Genet.* *17*, e1009368. 10.1371/journal.pgen.1009368.
 15. Kherreh, N., Cleary, S., and Seoighe, C. (2022). No evidence that HLA genotype influences the driver mutations that occur in cancer patients. *Cancer Immunol. Immunother.* *71*, 819–827. 10.1007/s00262-021-03028-w
 16. Castro, A., Ozturk, K., Pyke, R.M., Xian, S., Zanetti, M., and Carter, H. (2019). Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med. Genomics* *12*, 107. 10.1186/s12920-019-0544-1.
 17. Lakatos, E., Williams, M.J., Schenck, R.O., Cross, W.C.H., Househam, J., Zapata, L., Werner, B., Gatenbee, C., Robertson-Tessi, M., Barnes, C.P., et al. (2020). Evolutionary dynamics of neoantigens in growing tumors. *Nat. Genet.* *52*, 1057–1066. 10.1038/s41588-020-0687-1.
 18. Van den Eynden, J., Jiménez-Sánchez, A., Miller, M.L., and Larsson, E. (2019). Lack of detectable neoantigen depletion signals in the untreated cancer genome. *Nat. Genet.* *51*, 1741–1748. 10.1038/s41588-019-0532-6.
 19. Zapata, L., Caravagna, G., Williams, M.J., Lakatos, E., Abduljabbar, K., Werner, B., Chowell, D., James, C., Gourmet, L., Milite, S., et al. (2023). Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. *Nat. Genet.* *55*, 451–460. 10.1038/s41588-023-01313-1.
 20. Los Alamos National Security (2017). CTL/CD8+ Epitope Variants and Escape Mutations. https://www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html.
 21. Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* *8*, 33. 10.1186/s13073-016-0288-x.
 22. Keşmir, C., Nussbaum, A.K., Schild, H., Detours, V., and Brunak, S. (2002). Prediction of proteasome cleavage motifs by neural networks. *Protein Eng. Des. Sel.* *15*, 287–296. 10.1093/protein/15.4.287.
 23. Calis, J.J., de Boer, R.J., and Kesmir, C. (2012). Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. *PLoS Comput Biol* *8*, e1002412. 10.1371/journal.pcbi.1002412.
 24. Gartner, J.J., Parkhurst, M.R., Gros, A., Tran, E., Jafferji, M.S., Copeland, A., Hanada, K.-I., Zacharakis, N., Lalani, A., Krishna, S., et al. (2021). A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types. *Nat. Cancer* *2*, 563–574. 10.1038/s43018-021-00197-6.
 25. Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G.L., Eisenhaure, T.M., et al. (2017). Mass Spectrometry Profiling of HLA-Associated

- Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* 46, 315–326. 10.1016/j.immuni.2017.02.007.
26. Juncker, A.S., Larsen, M.V., Weinhold, N., Nielsen, M., Brunak, S., and Lund, O. (2009). Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. *PLoS One* 4, e7448. 10.1371/journal.pone.0007448.
 27. Fortier, M.H., Caron, E., Hardy, M.P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P. (2008). The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* 205, 595–610. 10.1084/jem.20071985.
 28. Garcia Alvarez, H.M., Koşaloğlu-Yalçın, Z., Peters, B., and Nielsen, M. (2022). The role of antigen expression in shaping the repertoire of HLA presented ligands. *iScience* 25, 104975. 10.1016/j.isci.2022.104975.
 29. Heemskerk, B., Kvistborg, P., and Schumacher, T.N. (2013). The cancer antigenome. *EMBO J* 32, 194–203. 10.1038/emboj.2012.333.
 30. Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 171, 1029–1041.e21. 10.1016/j.cell.2017.09.042.
 31. Patel, S.J., Sanjana, N.E., Kishton, R.J., Eidizadeh, A., Vodnala, S.K., Cam, M., Gartner, J.J., Jia, L., Steinberg, S.M., Yamamoto, T.N., et al. (2017). Identification of essential genes for cancer immunotherapy. *Nature* 548, 537–542. 10.1038/nature23477.
 32. Pearson, H., Thibault, P., Perreault, C., Pearson, H., Daouda, T., Granados, D.P., Durette, C., Bonneil, E., Courcelles, M., Rodenbrock, A., et al. (2016). MHC class I – associated peptides derive from selective regions of the human genome Find the latest version: The Journal of Clinical Investigation MHC class I – associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* 126, 4690–4701. 10.1172/JCI88590.defective.
 33. Capietto, A.-H., Hoshyar, R., and Delamarre, L. (2022). Sources of Cancer Neoantigens beyond Single-Nucleotide Variants. *Int. J. Mol. Sci.* 23, 10131. 10.3390/ijms231710131.
 34. Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., et al. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* 18, 1009–1021. 10.1016/S1470-2045(17)30516-8.
 35. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171, 1259–1271 e11. 10.1016/j.cell.2017.10.001.
 36. Bjerregaard, A.M., Nielsen, M., Jurtz, V., Barra, C.M., Hadrup, S.R., Szallasi, Z., and Eklund, A.C. (2017). An analysis of natural T cell responses to predicted tumor neoepitopes. *Front. Immunol.* 8, 1–9. 10.3389/fimmu.2017.01566.

-
37. Apps, R., Meng, Z., Del Prete, G.Q., Lifson, J.D., Zhou, M., and Carrington, M. (2015). Relative Expression Levels of the HLA Class-I Proteins in Normal and HIV-Infected Cells. *J. Immunol.* *194*, 3594–3600. 10.4049/jimmunol.1403234.
 38. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 15545–15550. 10.1073/pnas.0506580102.
 39. Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004). The Immunobiology of Cancer Immunosurveillance and Immunoediting. *Immunity* *21*, 137–148. 10.1016/j.immuni.2004.07.017.
 40. Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. *Science* *348*, 69–74. 10.1126/science.aaa4971.
 41. Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* *377*, 2500–2501. 10.1056/NEJMc1713444.
 42. Kalaora, S., Wolf, Y., Feferman, T., Barnea, E., Greenstein, E., Reshef, D., Tirosh, I., Reuben, A., Patkar, S., Levy, R., et al. (2018). Combined analysis of antigen presentation and T-cell recognition reveals restricted immune responses in melanoma. *Cancer Discov.* *8*, 1366–1375. 10.1158/2159-8290.CD-17-1418.
 43. Lang, F., Schrörs, B., Löwer, M., Türeci, Ö., and Sahin, U. (2022). Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat. Rev. Drug Discov.* *21*, 261–282. 10.1038/s41573-021-00387-y.
 44. Martínez-Jiménez, F., Priestley, P., Shale, C., Baber, J., Rozemuller, E., and Cuppen, E. (2023). Genetic immune escape landscape in primary and metastatic cancer. *Nat. Genet.*, 1–12. 10.1038/s41588-023-01367-1.
 45. Buckley, P.R., Lee, C.H., Ma, R., Woodhouse, I., Woo, J., Tsvetkov, V.O., Shcherbinin, D.S., Antanaviciute, A., Shughay, M., Rei, M., et al. (2022). Evaluating performance of existing computational models in predicting CD8+ T cell pathogenic epitopes and cancer neoantigens. *Brief. Bioinform.* *23*, bbac141. 10.1093/bib/bbac141.
 46. Oliveira, G., Stromhaug, K., Klaeger, S., Kula, T., Frederick, D.T., Le, P.M., Forman, J., Huang, T., Li, S., Zhang, W., et al. (2021). Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature* *596*, 119–125. 10.1038/s41586-021-03704-y.
 47. Cattaneo, C.M., Battaglia, T., Urbanus, J., Moravec, Z., Voogd, R., de Groot, R., Hartemink, K.J., Haanen, J.B.A.G., Voest, E.E., Schumacher, T.N., et al. (2023). Identification of patient-specific CD4+ and CD8+ T cell neoantigens through HLA-unbiased genetic screens. *Nat. Biotechnol.* 10.1038/s41587-022-01547-0.

48. Lin, X., George, J.T., Schafer, N.P., Ng Chau, K., Birnbaum, M.E., Clementi, C., Onuchic, J.N., and Levine, H. (2021). Rapid assessment of T-cell receptor specificity of the immune repertoire. *Nat. Comput. Sci.* 1, 362–373. 10.1038/s43588-021-00076-1.
49. Wang, L., and Lan, X. (2022). Rapid screening of TCR-pMHC interactions by the YAMTAD system. *Cell Discov.* 8, 30. 10.1038/s41421-022-00386-2.
50. Kula, T., Dezfoulian, M.H., Wang, C.I., Abdelfattah, N.S., Hartman, Z.C., Wucherpfennig, K.W., Lyerly, H.K., and Elledge, S.J. (2019). T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes. *Cell* 178, 1016-1028.e13. 10.1016/j.cell.2019.07.009.
51. Gejman, R.S., Jones, H.F., Klatt, M.G., Chang, A.Y., Oh, C.Y., Chandran, S.S., Korontsvit, T., Zakhaleva, V., Dao, T., Klebanoff, C.A., et al. (2020). Identification of the Targets of T-cell Receptor Therapeutic Agents and Cells by Use of a High-Throughput Genetic Platform. *Cancer Immunol. Res.* 8, 672–684. 10.1158/2326-6066.CIR-19-0745.

Methods

Datasets

HIV peptide data for pipeline prediction validation

We compiled a list of 81 HLA-A*02:01-restricted HIV T-cell epitopes by querying the Los Alamos National Security HIV Database (<https://hiv.lanl.gov>, December 2010) and PubMed. The obtained peptides were filtered according to the following criteria to create a list of ‘high confidence’ HLA-A*02:01-restricted and T-cell recognized epitopes: (1) epitope presentation was shown to be restricted to the HLA-A*02:01 allele; (2) T-cell reactivity against the epitope was reported in at least 3 patients/studies, and (3) evidence of endogenous processing of the epitope had been obtained (i.e., T-cell responses observed in vaccination or peptide loading studies were excluded). These criteria were met by 32 of 81 acquired peptides.

To assess all candidate HIV-1 epitopes, we acquired the assembled sequencing data of an HIV-1 isolate from the NCBI database (isolate 671-00T36; NCBI accession number AY423387¹) and considered it a reference HIV-1 genome. Partitioning this reference genome in all nonameric peptides, 3,094 candidate peptides were generated. Out of the 32 ‘high confidence’ HIV epitopes discussed above, 17 were not perfectly mappable to the reference genome (1-2 amino acid differences at most). To correct for this and allow cross-matching, the reference genome was adjusted to exactly match the mismatching peptide sequences (adjusted reference in Table S2). One out 17 epitopes remained completely unmappable to the reference genome and was thus excluded from further analysis (Table S2).

IEDB peptide data for pipeline validation

Peptides selected for T cell recognition proven using any methodology were downloaded from the Immune Epitope Database (<http://www.iedb.org>², 2018-12-10) from the ‘Assays’ section. Assays were filtered for: i) having four-digit HLA-typing (e.g., ‘HLA*B-27:05’), ii) having the targeted peptide be available and 9 amino acids of length in the (Antigen Description field) iii) having an entry for the Organism species name. Peptides were considered T-cell targetable if at least 2 tested subjects responded, combined over all the assays investigating a particular peptide. If this field was not available for any of the assays for a particular peptide, the peptide was considered T-cell targeted if at least one assay gave a positive result (i.e., response labelled as Positive, Positive-High, Positive-Intermediate or Positive-Low). Notice that these criteria are substantially less stringent than the ones employed for the focused HIV-set. This analysis was restricted to HLA-A*02:01 peptides as this allele was by far the best represented in the IEDB. Next, redundancy in the acquired T-cell targeted peptide sets per pathogen was removed by iterative sequence alignments of all unordered peptides

against all remaining peptides using the function `pairwiseAlignment` from the R Bioconductor package `Biostrings` (version 3.8). Matching scores above zero were interpreted as sequence similarity, resulting in removal of the second peptide of the pair from the peptide set. This way, the T-cell targeted peptides in the resulting list were all dissimilar from each other. Next, in order to predict proteasomal processing efficiency of candidate peptides, the amino acid context of the peptides in their source protein was required, but this information was not included for all peptides in the IEDB. To obtain this information, peptides were mapped to the reference proteomes of the viruses they were annotated to originate from using `phmmer` (version 3.2.1) with the `--max` and `--domtblout` flags, prioritizing matching reference sequences by their alignment length (the number of matching amino acids between query and target sequence, longer alignments preferred), the alignment discrepancy (when available, the difference between the annotated C-terminus and the inferred one, smaller is preferable), the source of the reference sequence (the manually curated SwissProt prioritized over the more exhaustive TrEMBL), the alignment's *e*-value and the query name, in that order. Query peptides with more than 3 mismatches between the source and query sequences were excluded from further analysis. The following viruses reference proteomes were used: Human gammaherpesvirus 4 (EBV, UP000007639), Human Immunodeficiency Virus 1 (HIV-1, AUP000002241), Alphapapillomavirus 9 (AUP000009104), Vaccinia virus (AUP000000344), Influenza-A virus (AUP000131152), Hepatitis-B virus (AUP000008591), Hepacivirus-C (AUP000000518), Human alpha-herpesvirus 1 (AUP000106517), Dengue virus (AUP000002500), Human betaherpesvirus-5 (AUP000000938). Transcriptome references were downloaded from the UniProt database by querying for the virus name and downloading all (possibly redundant) proteins (The UniProt, 2017). For non-perfectly mapping peptides, the most highly rated reference sequence was then modified to reflect the query sequence (i.e., peptide), such that the most representative processing score for the peptide could be computed.

TCGA data

The patient sample cohort, consisting of all tumor samples for which both DNA, RNA sequencing data was available, spans 5,585 patients from 30 tumor types (Table S1). TCGA data acquisition from the Broad Institute's Firehose and integration and preprocessing of data sources listed below was automated in R using functionality that is combined in the R package `firehosedownload`: <http://www.github.com/slagtermaarten/firehosedownload>.

TCGA somatic variant calls in MAF-format and RNA sequencing data were downloaded and harmonized from the 2015-08-21 release of the Broad TCGA genome data analysis center standard runs: <http://gdac.broadinstitute.org/runs/stddata>. As mutation data for the ESCA (oesophagus carcinoma) project is not part of this release, mutations for the ESCA project were obtained from the repository of mutations files curated by Cyriac Kandoth: <http://www.synapse.org/#!Synapse:syn1695396.13>. TCGA RNA sequencing data were downloaded from the Broad TCGA genome data analysis center 2015-11-01 release of the standard runs: <http://gdac.broadinstitute.org/runs/stddata>.

For projects where data from multiple sequencing platforms were available, Illumina HiSeq data was preferentially used. Raw read counts in RNAseq data were subsequently transformed to the transcripts per million (TPM) RNA abundance measure, using custom R functionality included in `firehosedownload` and using Ensembl75 (release of February 2014) gene length information.

TCGA HLA typing data

HLA typing based on DNA sequencing using OptiType³ was downloaded from the TCIA resource (www.tcia.at) provided by the Trajanoski laboratory.

Genome- and patient-level annotation

MMR status of tumor samples

The R package `MSIseq`⁴ was used to infer microsatellite instability status of all TCGA tumor samples in our cohort. Mutation annotation format files were obtained from TCGA as described above. Sequencing target region sizes were calculated for each sample from target enrichment design files used in the various TCGA projects (Supplementary Table S3). These data were subsequently used as input for the `MSIseq` classifier.

Annotation of antigen presentation capability and T cell sensitivity of tumor samples

To identify tumor samples that may be resistant to T cell attack, we analyzed samples for occurrence of non-synonymous mutations in any of the 515 non-HLA genes identified to potentially induce resistance to CD8⁺ T cell mediated killing⁵, or for occurrence of non-synonymous mutations in any HLA class I allele⁶.

PAM50 subtyping of breast cancer samples

We downloaded Level 3 RNA-seq data for the BRCA cohort from the TCGA Data Portal on 2015-06-25 and analyzed the expectation maximization normalized counts. Samples were PAM50-subtyped independently⁷ using the implementation in the `genefu` R package⁸ with the robust scaling option enabled.

Allele-specific HLA loss

We adapted the tool LOHHLA⁹, which allows for allele-specific detection of genomic aberration of HLA alleles, to make it more amenable to large scale application (code available on GitHub: <http://github.com/slagtermaarten/LOHHLA>). Most importantly, we included support for the reference genome GRCh38, made it compatible with single-end sequencing data and expanded it such that it can handle input bam files that are restricted to the HLA region of interest, rather than whole exome or whole genome bam files. The original version of LOHHLA compares whole-genome or whole-exome coverage mapped reads (bam formatted) between a tumor and matched normal sample, comparing read coverage normalized to the total number of mapped reads between the two samples for each HLA allele separately. To circumvent having to download the full bam files for thousands of patients just to obtain the total number of mapped reads per bam file, we inferred the total number of mapped reads from the file sizes of the complete bam files (accessible from the NIH GDC API). For this an ordinary linear regression model was used, which was fitted on file sizes in bytes and the number of mapped reads (as read out using samtools flagstat, 5th row of output) of 114 normal and tumor sample bam files for 60 randomly sampled patients (Table S5) for which we did download the entire bam file using gdc-client, deriving the expression: total mapped reads = $88370554 + 7461 [\text{reads/MB}] * \text{file size [MB]}$ (Figure S4.6A). This allowed us to use the TCGA GDC bam slicing API (<https://api.gdc.cancer.gov/slicing/view/>) to specifically download the HLA regions of chromosome 6 (29941260-29945884, 31353872-31357187 and 31268749-31272092 in GRCh38 coordinates) and use those as input for LOHHLA. Note that the full TCGA bam files that were obtained largely (~99.9%) consisted of mapped reads, obviating the need to correct for the presence of unmapped reads. In order to ensure that read pairs for which one of the read mates lied outside the annotated HLA genomic range were included in the analysis, the bam slicing download window was extended by 10^7 bp on both the 3' and 5' sides to 28941260-32357187. We fed the LOHHLA analyses purity and ploidy estimates obtained from ASCAT as described in the section 'Variant cellularity' below. The minCoverage argument to LOHHLA determines the minimally required amount of coverage in the normal sample in order for an SNP to be considered eligible to contribute to the copy number estimate of the gene it's positioned in. It was set to 0, after having tested the stability of the copy number estimates by titrating the minCoverage filter between 0 and a patient specific upper boundary computed as the median coverage of sites in the matched normal sample. While titrating the minCoverage threshold, we recorded the inferred copy number for each allele (i.e., $\text{HLA_type}\{1,2\}\text{copyNum_withBAF}$, which is computed as the median of the copy number estimates for individual loci pertaining to an allele, Figure S6B). Next, we computed the coefficient of variation (CoV, standard deviation/mean) over the copy number estimates to obtain donor, gene, and allele-specific estimates of the effect of the minCoverage threshold on the copy number estimates (Figure S6C). Alleles that showed a $\text{CoV} < .25$ were labelled as robustly estimable (90% of alleles, Figure S6D), resulting in 63% of evaluated patients to have robust estimates for all of their HLA-A, -B, and -C alleles (Figure S6E). Immunoediting analyses with the 'strict LOHHLA' attribute included only these patients; analyses with the 'lenient LOHHLA' attribute also included patients for

which not all alleles were robustly estimable. An overview of allele-specific copy number estimates for HLA-A, -B, and -C are displayed in Figure S6F. Alleles were deemed lost if the upper boundary of the copy number estimate's 95% confidence interval was below 0 and the estimate's minCoverage CoV was below .25.

Cytolytic score

Cytolytic score¹⁰, a transcriptomic proxy for T and NK-cell activity, was computed as the geometric mean of TPM-normalized expression estimates of the *PRF1* and *GZMA* genes.

Gene-level annotation

Gene essentiality

To infer the essentiality of genes to cell survival, we integrated work of the Sabatini and Brummelkamp laboratories. Wang et al.¹¹ screened the Jiyoye and Raji cell lines (Burkitt Lymphoma) and the K562 and KBM7 (CML) cell lines using CRISPR technology, and the KBM7 cell line using GeneTrap technology. Blomen et al.¹² screened the KBM7 cell line and its derivative, the HAP1 cell line, using GeneTrap technology. Blomen et al. provided binary class labels indicating essentiality for genes, whereas Wang et al. only provided raw read count data, presenting significance testing for only a subset of performed experiments.

In the Wang et al. CRISPR KBM7 data analysis, essential genes were considered to have a CRISPR-score (average \log_2 fold change in the abundance of sgRNAs targeting the gene) lower than -0.1 and an adjusted p -value below 0.05. Neither general nor cell line-specific criteria were included for the other three cell lines that Wang et al. screened using CRISPR technology. We elected to apply the same criteria to the other three cell lines screened with CRISPR technology by Wang et al. Similarly, no cut-off was proposed for the Wang et al. GeneTrap data set, but rather a correlation between the GeneTrap and CRISPR data was reported. Here, we used a minimum required amount of anti-sense inserts of 65 and set the required GeneTrap score to be lower than .45, in order to maximize the similarity in set cardinalities between the Sabatini-KBM7-GeneTrap and the Sabatini CRISPR/KBM7 derived gene set (1,875 and 1,878 genes, respectively). Having obtained seven partially overlapping lists of essential genes, any gene appearing at least once in any of the seven experiments was deemed essential, 10.6% of selected genes appeared in all lists. The list of genes is included in Table S4.

Peptide-level annotation

Similarity to self-repertoire

To determine whether predicted neoantigens were likely to be dissimilar enough from self to be recognized by the endogenous T-cell repertoire, we implemented a self-similarity classifier based on previously identified determinants of T cell similarity¹³. We compare candidate epitopes arising from somatic mutations to peptides from the human proteome predicted to be presented by the relevant HLA allele (e.g., HLA-A*02:01), restricting this comparison to the amino acids spanning positions 3 to 8 of nonameric peptides, as these are considered to be most important for T-cell recognition of peptide-HLA complexes¹³⁻¹⁵. Epitopes were deemed ‘dissimilar-from-self’ when one or more of the following criteria are met: (1) amino acid position 5 is mutated, (2) T-cell exposed region contains 3 or more mutations, (3) two mutations are clustered to one side of position 5 (i.e., in positions 2-3-4 or positions 6-7-8), and (4) a single amino acid substitution leads to large physicochemical changes on position 2, 3, 4, 6 or 7. The latter substitutions were defined as amino acid changes with an absolute covariance of < 0.05 in the PMBEC amino acid similarity matrix¹⁶. We implemented this algorithm in Rcpp (C++) and distributed it as part of the quickMHC R package (<http://www.github.com/slagtermaarten/quickMHC>).

Somatic variant annotation

Variant effect prediction

To determine the effects of the various classes of mutations found in tumor samples, we developed a Perl tool named VarContext. Canonical cDNA transcripts were obtained for genes containing mutations (single nucleotide variants and/or indels) by querying the Ensembl database (release 75, GRCh37). We applied each mutation affecting a particular gene and annotated its effect (silent, missense, insertion, deletion, frameshift, stop loss, stop gained), yielding the tumor transcript sequence. Transcripts which lost their stop codon as a consequence of the applied mutations were extended until the next in-frame stop codon was encountered. In contrast, transcripts gaining a premature termination codon (PTC) were analyzed for their potential of triggering nonsense-mediated decay (NMD; detailed description in ‘Transcript-level annotation’). Finally, the canonical and modified cDNA sequences were translated into amino acid sequence, resulting in the reference and tumor transcripts.

Variant oncogenicity and essentiality

The oncogenic potential of mutations was predicted using IntOGen mutations¹⁷ version ‘IntOGen Mutations Analysis 2.4.1-maintenance’ (<https://www.intogen.org/analysis>), specifically for all affected transcripts and stratified by sequencing project. TCGA maf files were converted to the required TSV-based input file format using a custom script and were analyzed in December 2015. Variants were further annotated with a driver gene score indicating the likelihood of being a driver gene based on the gene of their location¹⁸. To infer essentiality to cell survival of individual genes, we integrated the work of two laboratories^{11,12} into an overall essentiality score, with which somatic mutations were subsequently annotated (see ‘Gene essentiality’). Additionally, annotation of oncogenicity was performed by comparing gene, amino acid change and amino acid change position (determined using VarContext) of somatic mutations to the list of 1,018 recurrent oncogenic mutations compiled by Marty et al.¹⁹. The latter list was used for the ‘Marty’ variant selection setting.

Variant cellularity

Clonal antigens that are presented by all tumor cells in a lesion can reasonably be expected to have a larger contribution to tumor regression than subclonal antigens, and the observed inverse relationship between tumor heterogeneity and immunotherapy outcome provides indirect support for the superior value of clonal antigens as T-cell targets²⁰.

We estimated the cellularity of individual mutations from DNA sequencing read count information for a subset of 3,660 tumors in 18 tumor types, selected based on availability of the required data types. As a starting point for inferring the fraction of tumor cells that carry a mutation (the variant’s cellularity), we use the variant’s observed allelic frequency (VAF), the fraction of reads overlapping with the variant locus that carries the variant. The observed VAF of a somatic variant does not relate to its cellularity in a straightforward way as it is a compound measure of several factors: the proportion of contaminating normal cells, the number of allelic copies of the variant in each cell and its cellularity, plus uncharacterized sources of technical noise²¹. We employed two methods of inferring variant cellularity. The first is based on published maximum likelihood-based approaches^{20,22}, the second is a Bayesian hierarchical clustering of variants²¹ for potentially more robust and accurate cellularity estimates.

First, variants were annotated with the absolute copy number status of the genomic segments they are located in. Absolute copy number status was derived from ASCAT analyses of Affymetrix SNP6 profiles²³ and were obtained from the COSMIC resource <https://cancer.sanger.ac.uk/cosmic/download> on 2016-04-24. Absence of coverage in the ASCAT file was assumed to imply absence of local copy number aberrations as only small parts of the genome are covered in the SNP6 output. Additionally, tumor purity estimates, representing the percentage of tumor cells in the sample, were obtained from these ASCAT analyses. The intersection of patients eligible for neoantigen prediction

(patients for which DNA and RNA sequencing data was available) and those for which read count information and a COSMIC ASCAT analysis was available (required for cellularity estimates) consisted of 3689 TCGA patients, distributed over 13 sequencing projects (see Supplemental Table 1 for the exact list of included samples).

Maximum likelihood-approach for cellularity estimation Following Landau et al.²², we note that the observed number of reads consistent with the called mutation, N_a , is binomially distributed: $P(N_a) \sim \text{Bin}(N, AF_e(c))$, where N denotes the total number of reads covering the genomic locus of the variant and AF_e denotes the expected allele fraction of the variant under a particular fraction of cells carrying the mutation c (for cellularity), on which AF_e depends as follows:

$$AF_e = p \frac{c m}{(A_n(1-p) + A_t p)}$$

In which p denotes the tumor purity of the sample, i.e., the fraction of cells that are cancer cells, A_n and A_t denote the average amount of alleles in the normal and tumor populations, respectively, and m denotes the variant's multiplicity, i.e., the number of tumor alleles that carry the variant - an integral number smaller than or equal to A_t assumed to be equal across all tumor clones. The listed expression for AF_e can be understood as the fraction of the number of tumor cell alleles carrying the mutant allele and the total number of alleles, from both tumor and normal cells, at the somatic variant's genomic locus.

Two of the quantities on which AF_e depends are not directly observed: c and m . Thus, neither c nor m are unambiguously identifiable without knowledge or assumptions about the other. In the case where the VAF equals the fraction of alleles derived from tumor cells ($VAF = p * \frac{A_t}{p * A_t + (1-p) A_n}$), only one scenario is plausible: $c = 1$ and loss-of-heterozygosity (LOH) must have occurred at the variant loci in the cell giving rise to all sequenced cells ($m = A_t$). Ambiguity however arises for mutations located in genomic regions of copy number aberrations and for which the VAF does not unequivocally indicate m to equal CN_t (i.e., $VAF < 1$). This ambiguity is caused by the fact that a subclonal mutation (low cellularity) with high multiplicity could result in similar VAF -values as a clonal variant (high cellularity) with low multiplicity. Following the examples of Landau et al. and McGranahan et al.²⁴, implicit in the case of McGranahan et al., we assume the multiplicity of the variant to be unity when estimating cellularity, thereby running the risk of overestimating the c of somatic variants for which the multiplicity potentially exceeds one, i.e., variants located in an amplified segment. For some variants, it is certain that the multiplicity must have been greater than unity as $m = 1$ results in $VAF)AF_e$ for $c = 1$. For these, we iteratively increase m by 1 until VAF falls in the range of expected allelic frequencies (AF_e), stopping before m exceeds the major allele count. Having set the multiplicity m to a minimal value consistent with the observed data, we can proceed to compute the most likely cancer cell fraction c . Assuming a discrete uniform prior on c in the range $[0, 1]$, discretized in intervals of $1/1,000$, we compute the likelihood of each c under the

binomial model $P(c)$ and subsequently normalize the likelihoods by dividing them by their sum - the constant of proportionality of $P(c)$ - to obtain the posterior probability mass function of $P(c)$ ²². The mode of this distribution, i.e., the maximum likelihood estimate is taken as the somatic variant's cellularity, with the two boundaries centered on the mode and encompassing 95% of the distribution as the 95% credibility interval. Following McGranahan et al., we labelled mutations as clonal if their inferred c was greater than or equal to 0.95 and the upper boundary of the c 95% credibility interval was greater than or equal to .99.

PyClone for cellularity estimation In addition to the ML-based approach, we ran PyClone²¹ for all samples with read count information, ASCAT purity and absolute copy number estimates available and a median total read count on called variants exceeding 100 reads (simulations succeeded for 780 tumors, see Supplemental Table 1 for the list of included patients). We ran the binomial model for 10^5 Markov Chain Monte Carlo iterations, using a Markov Chain burn-in of a 1,000 iterations and Markov Chain thinning of 5.

Correspondence between the two methods We hypothesized that the largest structural difference between the two methods lies in the inference of allelic frequency, which is only of relevance for mutations located in genomic regions with copy number aberrations. In support of this, the two methods yield highly similar and correlated clonality estimates for those mutations that do not require an inference of allele multiplicity (Figures S2A and S2B). As this class forms the large majority of mutations (98.75% of somatic variants), further analyses were performed using the maximum likelihood-based approach, because of its more clearly defined dependency on sequencing depth. To provide a validation of the obtained cellularity estimates, we compared cellularity estimates between the most highly recurrent mutations within a given tumor type and the aggregate of non-recurrent mutations. Consistent with expectations, for the majority of tumor types (15 out of 16) the most highly recurrent mutation was predicted to be significantly more clonal than the remainder of the mutations observed in that tumor type (Figure S2C). As we deemed the correspondence between the two methods satisfactory, further analyses were based on the ML-based method for its relative simplicity, computational efficiency and relatively low requirements on read coverage.

Variant-specific expression

To estimate the expression of individual somatic variants, we required the RNASeq mapped reads from the TCGA GDC which are mapped to GRCh38.d1.vd1. We used the bam slicing tool in order to prevent having to download entire .bam files. We first converted the variant loci from the hg19 to GRCh38 coordinate system, using the `rtracklayer` and `Granges` functions out of the `rtracklayer` (version 1.42.2) and `GenomicRanges` (version 1.34.0) R Bioconductor packages in combination with a liftover file obtained from UCSC: <ftp://hgdownload.cse.ucsc.edu/goldenPath/hg38/>

`liftOver/hg19ToHg38.over.chain.gz`. These transformed coordinates were subsequently used to query the TCGA GDC bam slicing tool, using a random UUID in case multiple UUIDs were listed for RNASeq of tumor samples from the patient. Next, the number of reads consistent with the variant and reference alleles was tallied by running samtools mpileup (version 1.9) and custom R functionality.

Neoantigen prediction

Tumor transcripts were reconstructed from SNVs and indels in order to obtain a set of candidate tumor-specific neoantigens. Candidate peptides whose genomic sequences were affected by multiple mutations were modified to reflect the consequences of all variants (1.75% of all candidate peptides). SNVs and frameshifting insertion-deletions can introduce premature termination codons (PTCs), rendering the encoded transcripts prone to degradation by the nonsense-mediated decay (NMD) machinery. To account for this, we implemented an NMD-calling routine based on previously inferred characteristics of NMD-targeted transcripts²⁵. As expected, PTC-inducing variants that were classified as invoking NMD had significantly lower variant allele fractions (VAF) in RNA sequencing data than in DNA sequencing data, indicating degradation of PTC-bearing transcripts (Figure S1G). Hence, NMD-predicted transcripts were removed from further analysis.

Reference and tumor transcripts were used as inputs for Neolution, the in-house neoantigen prediction pipeline that annotates tumor-specific transcript derived peptides with four scores representing the various stages of antigen presentation, classifying those passing all four filters as MHC-binding peptides likely to yield an immune response. We assess the following steps: (i) RNA expression, (ii) proteasomal processing and transport into the endoplasmic reticulum, (iii) MHC binding, and (iv) dissimilarity from self-antigens. First, we determined whether genes encoding candidate neoantigens are expressed, excluding – unless indicated otherwise – all peptides for which the associated gene had an expression level surpassing a constant threshold (default: 0). Alternatively, we applied a threshold to library size-normalized read counts at the variant level as described in ‘Variant expression’. Second, we used netChop²⁶ to predict the likelihood of successful peptide processing by the proteasome and TAP transport (NetChop score $\geq .5$). Third, we used netMHCpan3.0²⁷ to predict HLA class I binding affinity, by default employing a percentile rank threshold of 1.9 (which corresponds to 255 nM for HLA-A*02:01). To ensure constant prediction precision across tumors, we elected to use one allele at a time rather than adapting the predictions alleles to the HLA haplotype of the patient. We selected HLA-A*02:01, -A*11:01, -B*07:02, -B*27:05 and -B*40:01, based on their prior determined in accuracy in predicting nonamer binding affinity²⁷ and for their functional diversity. To expedite (repeated) usage of peptide affinity predictions, affinity predictions for all candidate peptides and all HLA alleles encountered in the TCGA patient set were pre-cached in a PostgreSQL database (version 9.5, querying code available in the R package quickMHC). Finally, we determined whether T cell recognition is likely to be prevented by self tolerance. As the majority of mutated antigens derive from single nucleotide variants and, by consequence, are highly similar to their wildtype

counterparts, it may be important to exclude candidate peptides that are too similar to the self-ligandome, as thymic negative selection eliminates T cell clonotypes reactive to antigens from these peptides. To this end, we implemented a ‘self-similarity’ filter^{13,14} that compares each candidate epitope to a reference list of self-epitopes predicted from the complete human proteome (obtained from UniProt, 2016-10). Candidate peptides were retained when deemed sufficiently different from self, according to the criteria outlined in ‘Similarity to self-repertoire’.

Antigen prediction pipeline validation

To evaluate the prediction performance of Neolution when performing predictions either (i) solely based on predicted HLA affinity, or (ii) also incorporating proteasomal processing predictions, or (iii) additionally excluding peptides with similarity to human sequences, a curated list of ‘high confidence’ HIV epitopes (see ‘HIV peptide data for pipeline prediction validation’ and Supplemental Table S2) and ‘lower confidence’ lists of peptides from the IEDB (see ‘IEDB peptide data for pipeline validation’) was processed using it. From these sets of predictions, prediction precision (PPV), sensitivity (TPR) and false positive rate (FPR) were computed using custom R code. To compute 80% confidence intervals (Cis) around PPV-estimates, we employed stratified bootstrapping to ensure constant prevalence of T-cell targeted peptides (and hence their presence) across bootstrapped samples. After quality metrics were computed for bootstrapped samples, the 10th and 90th percentiles of these metrics were taken as the upper and lower boundaries of the reported 80% Cis.

Somatic variant and neoantigen load tallying

In order to compare the propensity of different classes c of a particular classification C (e.g., DNA damage types) to yield HLA-A*02:01-antigens, we needed to tally the load of mutations and neoantigens for each class c in classification C . The majority of mutations can be unambiguously annotated as deriving from a single class c (e.g., ‘missense mutation’ in the transcript effect classification). However, some somatic mutations overlapping with the genomic loci of multiple genomic features (i.e., transcripts and genes) have different effects on these genomic features. To allow variants to belong to multiple classes c of a classification C during the tallying of variants, each variant was partially assigned to a class c based on the fraction of transcripts affected by the variant for which the effect of the mutation can be classified as c . This way, each variant potentially distributes its contribution among multiple classes, but its total contribution never exceeds unity (1). The total variant load of class c in a genome g , $N_{g,c}^v$, then becomes

$$N_{g,c}^v = \sum_{v \in V_g} \frac{|\{t \in T_v : S(t^v) = c\}|}{|T_v|}$$

where V_g denotes the complete set of variants in genome g , T_v the set of (Ensembl reference) transcripts affected by variant v , the S operator returns the variant effect classification of v on t and $|T_v|$ represents the number of transcripts affected by v (i.e., the cardinality of T_v).

A similar consideration of mixed class membership is made during the tallying of neoantigens that can be contributed by multiple somatic variants of potentially different classes. When tallying the neoantigen loads by different classes c of classification C , an epitope derived from a transcript t was taken to result from c proportional to the fraction of total variants overlapping with t classified as c with respect to t . This nuance is necessary as variants may have different effects on different transcripts. Combined, the peptide load contributed by somatic variant class c in genome g , $N_{g,c}^p$, becomes

$$N_{g,c}^p = \sum_{p \in P_g} \frac{|\{v \in V_p : S(v) = c\}|}{|V_p|}$$

where P_g denotes the complete set of peptides passing all filtering steps for genome g , V_p the set of variants contributing to peptide p , the S operator returns the class of its argument and $|V_p|$ represents the number of variants contributing to peptide p .

The ‘two group’ strategy for testing neoantigen depletion across unpaired samples

As immune pressure against mutations associated with HLA-A*02:01-peptides is not expected in patients that lack the HLA-A*02:01-allele and also any other alleles with similar binding profiles, we could compare the HLA-A*02:01-yield rates in HLA-A*02:01-positive patients (test set) with the yield rates in HLA-A*02:01-like negative patients (reference set). We did this both in a patient specific manner, comparing the distributions of patient specific yield rates against each other, and by aggregating all mutations and neoantigens into a meta-patient before comparing these point estimates. In the former case, distributions were compared using a two-sided Wilcoxon rank sum test implemented in the R base package (`wilcox.test`). In the latter case, proportions of neoantigens over total mutations were compared using a chi-squared test as implemented in R base package (`prop.test`). Effect sizes shown by color in Figures S3C & S3D are \log_2 differences, computed by logging the medians of the yield rates in the test and reference sets and subtracting them from each other. All FDR-multiple testing corrections were done per analysis (i.e., columns in Figure S3D) with Benjamini-Hochberg’s procedure²⁸ using the `p.adjust` function in R.

Detection power analysis

As mutational load and sample count both influenced the statistical power to detect a signal of epitope loss in our meta-patient immune editing analyses, the limits of detectable immune editing varied between sequencing projects. To determine these detection limits, we first normalized the observed yield rate in the test set to that observed in the reference set, equalizing any potentially pre-existing neoantigen yield rate imbalances between the test and reference sets. This way, we would measure the required rather than additionally required epitope loss for statistical significance. Next, we set out to determine the required immune pressure (IP), defined as the fraction of binding peptides that are lost compared to the reference (i.e., the relative yield rate decrease in test set patients) to reach statistical significance. As only true positive predicted binding peptides are immunologically visible and hence targetable, the prediction precision (PPV , fraction of true positive and all predicted peptides) downscales the IP such that the product of the two determines the actually observed immune strength, IP_o : $IP_o = IP \times PPV$. As such, we continuously lowered the neoantigen yield rate in the test set by a factor $1 - IP_o$, increasing IP_o from 0 to 1 in increments of .01, while testing for statistical significance using a chi-squared test for equal proportions (prop.test as implemented in R). For each combination of a neoantigen prediction PPV type, the required immune strength to reach this statistically significant effect size was recorded.

Using the silent mutational load to predict expected neoantigen load and scan for immune editing

Rooney et al.¹⁰ compared the observed neoantigen load to an expected neoantigen load computed from the silent mutational load, which can be assumed not to be penalized by T-cell pressure as synonymous mutations, unlike non-synonymous mutations, do not form neoantigens. The silent mutational load was subsequently used to estimate exposure to DNA damaging processes and thereby infer the number of (neoantigenic) non-synonymous mutations expected in the case of no selection pressure during tumor outgrowth. Assuming that the ratios between (1) synonymous mutations and non-synonymous mutational loads, and (2) non-synonymous mutations and neoantigen counts are on average equal between tumors, it is possible to compute the expected or predicted number of neoantigens from the silent mutational load and two sequentially applied conversion factors. The expected non-synonymous mutational load NS_p is estimated from the observed silent mutational load by multiplying it with the globally estimated conversion factor $c_{S \Rightarrow NS}$. Next, NS_p is used to compute the expected number of neoantigens E_p by multiplying it with the globally estimated conversion factor $c_{NS \Rightarrow E}$. We finally end up with R , the ratio between observed and predicted neoantigens for an individual sample

$$R = \frac{E_o}{E_p} = \frac{E_o}{S c_{S \Rightarrow NS} c_{NS \Rightarrow E}}$$

which relates individual samples to the remainder of the cohort which was used to compute the conversion factors.

To model the variable prevalence of various mutational processes operative in the genomes under analysis, Rooney et al. performed this analysis in a mutational spectrum specific fashion, i.e., by taking into consideration the nucleotides directly on the 5' and 3' sides of the mutated basepair (e.g., AC>TG, a C>T mutation flanked by an A and G). Following Rooney et al., conversion factors are here computed specifically for each of the 192 possible spectra. As such, R becomes the ratio of the observed neoantigens and the sum of the spectrum specific expected neoantigen loads.

There are a number of differences between the neoantigen prediction strategy employed here and in Rooney et al., precluding a direct copy of their methodology. First, the prediction pipeline employed by Rooney et al. does not account for indels and for interactions between mutations, i.e., potential interactions between mutations in the form of co-occurrence of mutations in a nonamer-spanning genomic sequence. To eliminate this source of possible discrepancy, the NS and S input data for this analysis are restricted to missense mutations. Analogously, peptides resulting from non-missense mutations are excluded. This means that a neoantigen yielded by for instance a missense mutation and an indel will not increase the missense mutation tally by one half but instead is excluded – it would not have been included in the neoantigen predictions used by Rooney et al. Second, Rooney et al. performed neoantigen predictions for patient-matched HLA types whereas we elected to use HLA-A*02:01-predictions for all analyzed donors because of the superior binding affinity predictions as compared to many other alleles. To further harmonize our methodologies, we excluded all variants (and associated neoantigens) located in driver genes, as these were also excluded by Rooney et al. (driver genes identified using MutSig, Table S6A from Rooney et al.). We performed these analyses on i) either only clonal mutations or all mutations, ii) in an either mutational context specific fashion or not for more robust conversion factor estimates (through including more events per conversion factor), and iii) using our 4-filtering neoantigen prediction pipeline or using predicted HLA affinity only, the latter to be more consistent with Rooney et al.. Conversion factors were recomputed per analysis on all included samples (i.e., pan-cancer).

The ‘continuous’ detection strategy -correlating neoantigen yield rates to HLA presentation scores across patients

Similar to the discrete-group approach to immunoediting testing, the fundamental idea here is to compare neoantigen yield rates for a particular four-digit HLA class I allele (hereafter called focus allele, e.g., HLA-A*02:01) between tumor samples containing HLA alleles that differ in their ability to present peptides that can be presented by this focus allele, but now analyzed on a continuous scale. In case depletion of neoantigenic non-synonymous mutations does occur, tumor samples that contain HLA alleles that show a large degree of overlap in binding pattern with that of the focus allele (i.e., that have a high presentation score, h) should show less neoantigens per mutation (low neoantigen

yield rate, r) on average, as compared to tumor samples that are poor at presenting these peptides (low h). We assess this difference in r between tumor samples that are high and low in h using linear regression.

HLA presentation corroboration

Overlap in HLA presentation (HLA corroboration score, Figure 2A) was assessed through overlap in predicted binding affinity over all 10,072,577 candidate nonameric neo-peptides that we processed for our patient cohort and for all 227 identified HLA class I alleles, in order to select a subset of all 20^9 potential nonamers that is representative of the human antigenome. The asymmetric HLA corroboration score for alleles I to allele j was computed as the fraction of peptides binding to j that are also predicted to bind I and thus falls in the range $[0, 1]$: $\frac{|P_i \cap P_j|}{|P_j|}$, where P_i is defined as the set of peptides presented by HLA allele i . Binding peptides were defined as having a NetMHCpan3.0 predicted percentile rank ≤ 1.9 (which corresponds to 255 nM binding affinity for HLA-A*02:01). Three other thresholds for the definition of binding peptides (1, 3 and 4, corresponding to 100, 500 and 1,000 nM binding affinity for HLA-A*02:01) did not substantially alter our results (data not shown).

Computation of tumor sample and HLA allele-specific HLA presentation scores

For the combination of a the HLA class I repertoire of an individual tumor sample and a particular focus allele for which immunoediting analysis was performed, the HLA presentation score (h) was defined as the fraction of peptides that is presentable by the focus allele that is also presentable by one or more of sample's HLA alleles, i.e., $h = \frac{|\{p : p \in P_f \text{ and } p \in P_r\}|}{|P_f|}$, in which P_f is the set of peptides presentable by the focus allele and P_r is the set of peptides presentable by the sample's class I HLA allele repertoire.

For analyses where allelic loss of HLA alleles (inferred using LOHHLA, see section 'Annotation of allele-specific HLA loss') is integrated into h_i , alleles identified to have been lost were excluded from A before computation. Allele amplification was ignored. Where the 'Presentation score' setting was set to 'HLA A, B' (rather than 'HLA A, B, C'), C-alleles were excluded, regardless of other settings.

Mass spec validation of HLA presentation score

To evaluate the validity of the sample-specific HLA presentation scores, we reasoned that mass spec-eluted HLA ligandomes should be enriched for peptides associated with HLA-alleles for which the corresponding patient has a high presentation score. To test this, we obtained mass spectrometry data of HLA-eluted peptides from 26 samples²⁹ through the SysMHC atlas (project ID: SYSMHC00010). We inferred the 4 HLA-A and -B alleles of each sample's corresponding HLA repertoire by consid-

ering the HLA alleles to which peptides had been assigned by the original investigators, assuming homozygosity and two allelic copies when only one HLA-A or B-allele was reported in total. Next, we compiled sample specific reference proteomes by aggregating all source proteins with which the observed peptides were annotated by the original investigators, so as to work around the unavailability of matched RNAseq data for these samples. We then assessed which peptides were expected to be present in the HLA ligandome of each combination of a sample and a focus allele by extracting all 9-mers from these proteins, using the UniProt human reference proteome (UP000005640_9606) as a reference, and annotating these peptides with a predicted affinity for the focus allele. The minority (<.1%) of peptides containing either a U (selenocysteine) or X (unknown amino acid) were excluded a priori. Having filtered these peptides, with passing peptides having an HLA affinity percentile rank ≤ 1.9 , we assessed which fraction of filter-passing peptides was observed in the experimental data to arrive at a quantity we interpreted here as prediction precision. Such predictions precisions were computed for all combinations of all 26 samples and 226 focus alleles (all HLA-A and -B alleles detected in the TCGA patients we analyzed).

A continuous approach to testing for immunoediting

We modeled the mean rate at which mutations yield neoantigens for a particular tumor type r and how that rate is modulated by the HLA presentation score (h) for the focus allele of the analysis. We thus define:

$$N_i^p \sim \beta_0 + N_i^v r_i + \varepsilon$$

$$r_i = (1 + \beta_h h_i) \beta_r$$

where N_i^p , N_i^v are the neoantigen and mutational load of the patient indexed by i , respectively, and r_i is the patient specific neoantigen yield rate, consisting of a global neoantigen yield rate β_r , shared across patients, and a second term that describes the degree to which the HLA presentation capability (presentation score h_i) shifts this global yield rate in a patient-specific manner. This second term consists of the patient-specific presentation score h_i and β_h , the fractional degree to which r_i is modulated by a unit increase in h_i . Finally, β_0 is an intercept term (the predicted number of neoantigens when $N^p = 0$ and $h = 0$) and ε represents the model residuals (i.e., $\widehat{N}_i^p - N_i^p$).

Integrating these two expressions and removing the coefficients, we get the following R model formula to be used in a R (g)lm:

$$N^p \sim 1 + N^v + h: N^v$$

Such that the first coefficient represents an intercept term, the second coefficient represents β_r and the third β_h .

A subanalysis was deemed reliably estimable if the p-value associated with β_r (i.e., the base r , the expected neoantigen yield rate for patients with $h = 0$) was < 0.05 . In cases of too few samples or the absence of a (linear) pattern in the data, this did not happen.

We were ultimately interested in the mean fractional difference Δ (delta) in neoantigen yield rate between patients with full HLA-presentation capability for the focus allele in question ($h = 1$) and those with no capability whatsoever ($h = 0$), which boils down to the following expression:

$$\Delta = \frac{(\beta_0 + \beta_r N^v + \beta_p N^v) - (\beta_0 + \beta_r N^v)}{\beta_0 + \beta_r N^v} = \frac{\beta_h}{\beta_0 / N^v + \beta_r}$$

in which N^v is the median number of mutational events across patients. The uncertainty in Δ was estimated using Monte Carlo simulation. We generated 10^4 random samples from the multivariate normal distribution of the regression coefficients (mean vectors from `coef` and covariance matrix from `vcov` methods of the R fit objects) using the `mvrnorm` function in the R MASS package (version 7.3.54). We computed Δ for each of these simulations and report the 10th and 90th percentiles of the resulting distribution (Figure 2D). Both N^v and N^p were transformed with $f(x) = \log_{10}(x + 1)$ prior to regression analysis.

Permutation testing Δ for significance

We employed permutation testing to assess the statistical significance of observed Δ values. Specifically, the presentation score for each sub-analysis was permuted 250 times and the resulting Δ was computed for each of those permutations. Any downward or upward bias resulting from distributional imbalances and other potential violations of linear model assumptions will be accounted for using this approach. We assessed the fraction of permutation Δ values that was larger than the observed Δ , a quantity we refer to as q for quantile. Small values of q (< 0.05) are indicative of a substantial deviation from the permutation null distribution and would be consistent with immunoediting. Large values (0.95) would likewise be consistent with neoantigen enrichment.

Detecting parameterizations associated with editing

Aggregating all sub-analyses, each a unique combination of all the potential discrete settings to make, resulted in a table with ‘ID’ columns, one for each of the configuration categorical variables, and columns for the resulting observed statistics, Δ and q . As we wanted to understand which settings increase the likelihood of observing negative Δ values, we sorted this table for either one of these statistics and univariately tested all ‘ID’ variables independently for enrichment. We repurposed fast

preranked gene set enrichment analysis testing here, as implemented in the fgsea R package (package version 1.18.0). For each level of a categorical variable (e.g., 'A*02:01' when considering the focus allele) occurring in at least 100 sub-analyses, we ran `fgsea::calcGseaStat` with `scoreType = 'std'` on the rankings of the sub-analyses and with all the positions of the level in question as the argument to `selectedStats`. The `calcGseaStat` function computes a running-sum statistic and returns the location where that statistic deviates most strongly from that expected under random shuffling of the level: the enrichment score (ES). We report the mean-centered version of this statistic, for each setting separately. Mean values were computed over all enrichment scores and deducted from the original ES.

Data and software availability

The TCGA data download is encapsulated in the R package `firehosedownload` (<http://www.github.com/slagtermaarten/firehosedownload>). Neoantigen predictions are performed using our custom software packages `VarContext` (<http://www.github.com/schumacherlab/varcontext>, projecting mutations on transcript RNA sequences to compute candidate neoantigenic peptides), `neolution-prep` & `neolution-live` (<http://www.github.com/schumacherlab/neolution-prep> and <http://www.github.com/schumacherlab/neolution-live>, for annotation and filtering of candidate peptides), supported by the `quickMHC` package located at <http://www.github.com/slagtermaarten/quickMHC>.

Neoantigen tallying is further wrapped in the `fasanalysis` (an acronym for foreign antigen space analysis) package (<http://www.github.com/slagtermaarten/fasanalysis>). Immunoediting analysis code, running the analyses and generating the figures presented in this work, is available at <http://www.github.com/slagtermaarten/immunoediting>. With the exception of `VarContext`, which is implemented in Perl, all functionality was implemented in R.

Bibliography

1. Geels, M.J., Cornelissen, M., Schuitemaker, H., Anderson, K., Kwa, D., Maas, J., Dekker, J.T., Baan, E., Zorgdrager, F., van den Burg, R., et al. (2003). Identification of sequential viral escape mutants associated with altered T-cell responses in a human immunodeficiency virus type 1-infected individual. *J Virol* 77, 12430–12440.
2. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* 47, D339–D343. 10.1093/nar/gky1006.

-
3. Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 30, 3310–3316. 10.1093/bioinformatics/btu548.
 4. Huang, M.N., McPherson, J.R., Cutcutache, I., Teh, B.T., Tan, P., and Rozen, S.G. (2015). MSIsq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci Rep* 5, 13321. 10.1038/srep13321.
 5. Patel, S.J., Sanjana, N.E., Kishton, R.J., Eidizadeh, A., Vodnala, S.K., Cam, M., Gartner, J.J., Jia, L., Steinberg, S.M., Yamamoto, T.N., et al. (2017). Identification of essential genes for cancer immunotherapy. *Nature* 548, 537–542. 10.1038/nature23477.
 6. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152–1158. 10.1038/nbt.3344.
 7. Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27, 1160–1167. 10.1200/JCO.2008.18.1370.
 8. Gendoo, D.M., Ratanasirigulchai, N., Schroder, M.S., Pare, L., Parker, J.S., Prat, A., and Haibe-Kains, B. (2016). Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* 32, 1097–1099. 10.1093/bioinformatics/btv693.
 9. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171, 1259–1271 e11. 10.1016/j.cell.2017.10.001.
 10. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. 10.1016/j.cell.2014.12.033.
 11. Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. *Science* 350, 1096–1101. 10.1126/science.aac7041.
 12. Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096. 10.1126/science.aac7557.
 13. Calis, J.J.A., de Boer, R.J., and Keşmir, C. (2012). Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. *PLoS Comput. Biol.* 8. 10.1371/journal.pcbi.1002412.
 14. Calis, J.J., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Kesmir, C., and Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* 9, e1003266. 10.1371/journal.pcbi.1003266.

15. Kessels, H.W., de Visser, K.E., Tirion, F.H., Coccoris, M., Kruisbeek, A.M., and Schumacher, T.N. (2004). The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. *J Immunol* *172*, 2324–2331.
16. Kim, Y., Sidney, J., Pinilla, C., Sette, A., and Peters, B. (2009). Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics* *10*, 394. 10.1186/1471-2105-10-394.
17. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A., and Lopez-Bigas, N. (2013). IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* *10*, 1081–1082. 10.1038/nmeth.2642.
18. Rubio-Perez, C., Tamborero, D., Schroeder, M.P., Antolin, A.A., Deu-Pons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2015). In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* *27*, 382–396. 10.1016/j.ccell.2015.02.007.
19. Marty, R., Kaabinejadian, S., Rossell, D., Slifker, M.J., van de Haar, J., Engin, H.B., de Prisco, N., Ideker, T., Hildebrand, W.H., Font-Burgada, J., et al. (2017). MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* *171*, 1272–1283 e15. 10.1016/j.cell.2017.09.050.
20. McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* *351*, 1463–1469. 10.1126/science.aaf1490.
21. Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Cote, A., and Shah, S.P. (2014). PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* *11*, 396–398. 10.1038/nmeth.2883.
22. Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* *152*, 714–726. 10.1016/j.cell.2013.01.019.
23. Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* *107*, 16910–16915. 10.1073/pnas.1009843107.
24. McGranahan, N., Favero, F., de Bruin, E.C., Birkbak, N.J., Szallasi, Z., and Swanton, C. (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* *7*, 283ra54. 10.1126/scitranslmed.aaa1408.
25. Lindeboom, R.G., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* *48*, 1112–1118. 10.1038/ng.3664.

-
26. Keşmir, C., Nussbaum, A.K., Schild, H., Detours, V., and Brunak, S. (2002). Prediction of proteasome cleavage motifs by neural networks. *Protein Eng. Des. Sel.* 15, 287–296. 10.1093/protein/15.4.287.
 27. Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 8, 33. 10.1186/s13073-016-0288-x.
 28. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.
 29. Pearson, H., Thibault, P., Perreault, C., Pearson, H., Daouda, T., Granados, D.P., Durette, C., Bonneil, E., Courcelles, M., Rodenbrock, A., et al. (2016). MHC class I – associated peptides derive from selective regions of the human genome Find the latest version: The Journal of Clinical Investigation MHC class I – associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* 126, 4690–4701. 10.1172/JCI88590.defective.

A

TCGA

DNA RNA

Determine candidate peptides
Predict NMD

Neolution
RNA expression, proteasomal processing,
HLA affinity, self-similarity

Testing
neo-antigen
depletion

Estimate of
prediction
performance

B

Prediction precision

Binding affinity threshold [nM]

Affinity Affinity Prot. processing Affinity Prot. processing Self-similarity

C

Precision

Sensitivity

Max precision (0.54) at APR \leq 40
Max F1 (0.49) at APR \leq 220

100 nM 500 nM 1000 nM

D

Sensitivity

False positive rate

E

Precision

Sensitivity

Max precision (0.06) at APR \leq 10
Max F1 (0.09) at APR \leq 10
116/169381 (6.85×10^{-4}) P

100 nM 500 nM 1000 nM

F

Sensitivity

False positive rate

G

Bladder

Melanoma

Transcript NMD labeling

no NMD

NMD

RNA / DNA VAF ratio

$p < 2 \times 10^{-16}$

$p < 2 \times 10^{-16}$

Figure S4.1: Prediction performance for (neo-)antigens and NMD-filter

A Schematic overview of data flow in this work.

B Prediction precision as a function of the HLA-affinity filter threshold in the curated HIV epitope set for HLA-A*02:01. Shaded areas denote 80% confidence intervals acquired using stratified bootstrapping.

C Parametric plot showing the trade-off between prediction precision (true binders amongst predicted binders) and prediction sensitivity (predicted true binders of all true binders) as a function of the HLA-affinity filter threshold in the curated HIV epitope set for HLA-A*02:01. Increasing the stringency of the affinity threshold increases precision with a trade-off in sensitivity. Dashed red line indicates the 255 nM affinity threshold used for further analyses.

D As in **A**, but plotting the trade-off between sensitivity and drop-out (FPR). Inset: zoom in on the horizontal range of [0, 0.03] of the ROC curve.

E As in **A**, but for the aggregate of all 10 viral genomes in IEDB (Methods).

F As in **B**, but for the aggregate of all 10 viral genomes in IEDB.

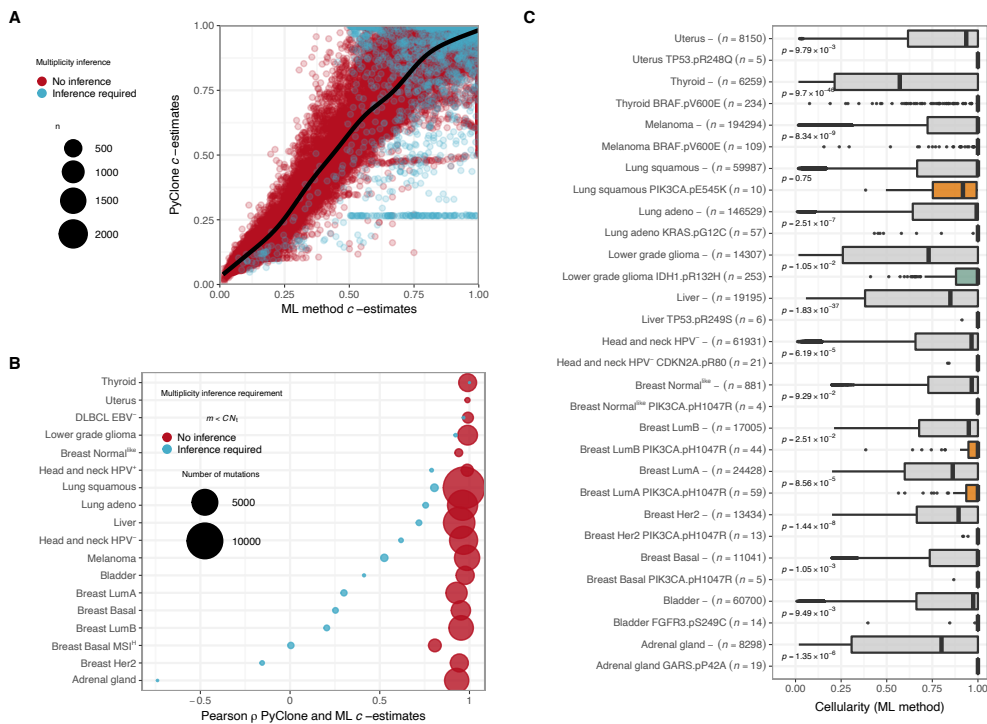


Figure S4.2: Validation of mutation clonality estimates.

A Correlation of cellularity estimates obtained by PyClone and maximum likelihood method. Color indicates whether a mutation is located in an amplified region and thus whether multiplicity inference is required. 98.75% of mutations do not strictly require multiplicity inference (blue dots).

B Pearson correlation between maximum likelihood and PyClone estimates of cellularity for each tumor type individually. Correlations are high for mutations where the multiplicity is certain but less for mutations for which multiplicity estimates are required (STAR methods). Dot size reflects the number of mutations on which the correlation coefficient is based.

C The mostly highly recurrent somatic mutations per cancer types, suggestive of a role in oncogenesis, are significantly more clonal than the aggregate of all non-recurrent mutations (indicated by a hyphen, grey boxplots) for 12/13 evaluated tumor types. Mutations were grouped based on their effects at the amino acid level. p -

values reflect one-sided Wilcoxon rank sum tests between non-recurrent and the most highly recurrent (across included tumor samples) mutations. Esophagus cancer and DLBCL were left out of this comparison as the most recurrent mutations were positioned in non-oncogenes for these tumor types (*CEP170* and *MUC6*, respectively).

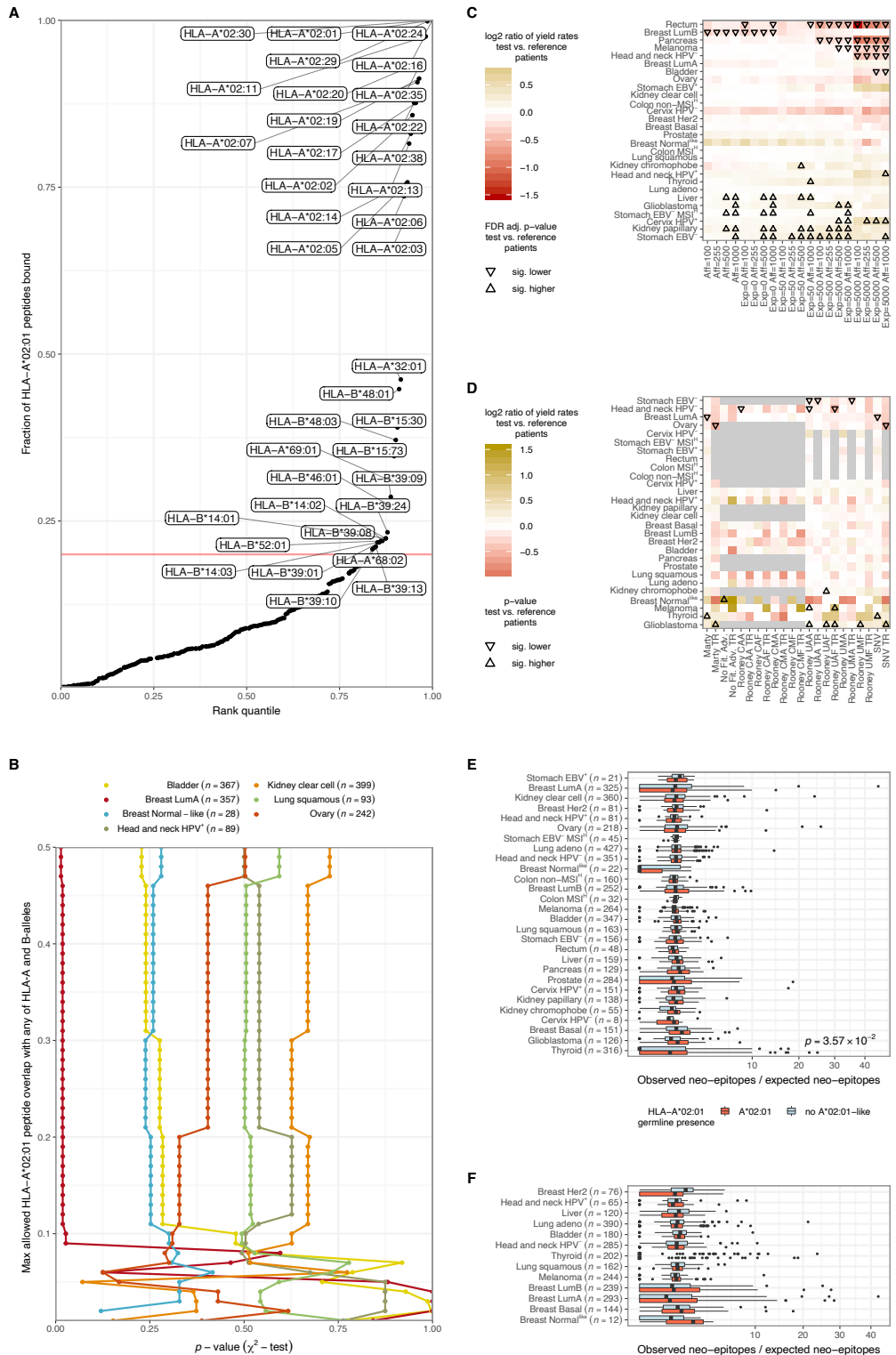


Figure S4.3: Extended analysis of genomic neoantigen depletion

A Fraction of predicted HLA-A*02:01 neoantigens that is predicted to also be presented by any of the other detected HLA class I A and B alleles. Unless indicated otherwise (see B), HLA alleles predicted to present more than 20% of predicted HLA-A*02:01-neoantigens (as indicated by the red line) were labelled as 'HLA-A*02:01-like' in subsequent analyses, and samples expressing any of these alleles were excluded from the HLA-A*02:01-like negative reference group.

B The relationship between 'the HLA-A*02:01-similarity threshold for patient inclusion in the HLA-A*02:01-like-negative reference set and the observed p -values in a meta-patient analysis of yield rate differences between the test samples and reference set samples is depicted per tumor type. To avoid over-plotting, seven randomly selected tumor types are shown. Note that an increase in stringency does not result in a systematic increase in statistical significance when comparing test and reference groups.

C Effect of different neoantigen prediction pipeline parameterizations on the detection of neoantigen depletion. Yield-rate differences in meta-patient analyses (see Figure 4.3A) are depicted. Triangles reflect comparisons for which FDR-adjusted p -values from two-sided Wilcoxon rank-sum tests were below 0.05. Upward and downward pointing triangles would point to genomic neoantigen depletion and to negative genomic neoantigen depletion, respectively.

D Effect of different modifications of neoantigen depletion analysis strategies on the detection of neoantigen depletion, while keeping neoantigen prediction parameters at the default settings. In the 'Marty'-analyses, analysis is restricted to recurring oncogenic mutations as identified by Marty et al. (2017). Also shown are analyses in which mutations that can confer a fitness advantage are excluded ('No Fit Adv.'), as in Figure 4.3D. In addition, 8 versions of analyses inspired by Rooney et al. (2015) are included. In this analysis, all permutations of three core settings were tested, denoted by three letters in the following format: '(C|U)(M|A)(F|H)'. Position 1 of the analysis identifier denotes the exclusion of variants based on clonality: (U)naware (no exclusion of subclonal mutations) vs. (C)lonal. Position 2 denotes the computation of conversion factors: (M)utational context specific vs. (A)specific. Position 3 denotes the neoantigen prediction pipeline filtering parameter: (F)ull 4-filter pipeline vs. HLA affinity filter (H) only. Finally, the SNV-restricted analysis also presented in Figure 4.3A is shown. Of all analyses, a T cell-resistance (TR) version is additionally shown, in which tumor samples that have at least one non-synonymous mutation in one of the 515 genes labelled as providing resistance to T-cell pressure³¹ are excluded. Unadjusted p -values from two-sided Wilcoxon rank-sum tests are shown by triangles in case of significance ($p < 0.05$). Upward and downward pointing triangles would point to genomic neoantigen depletion and to negative genomic neoantigen depletion, respectively. Missing analyses due to missing input data are indicated by a gray fill.

E Employing a mutation loss detection strategy inspired by Rooney et al. (2015), in which the expected neoantigen load is computed based on the synonymous mutation load and contrasted against the observed (predicted) neoantigen load.

F As in E, but also restricting the analysis to clonal mutations.

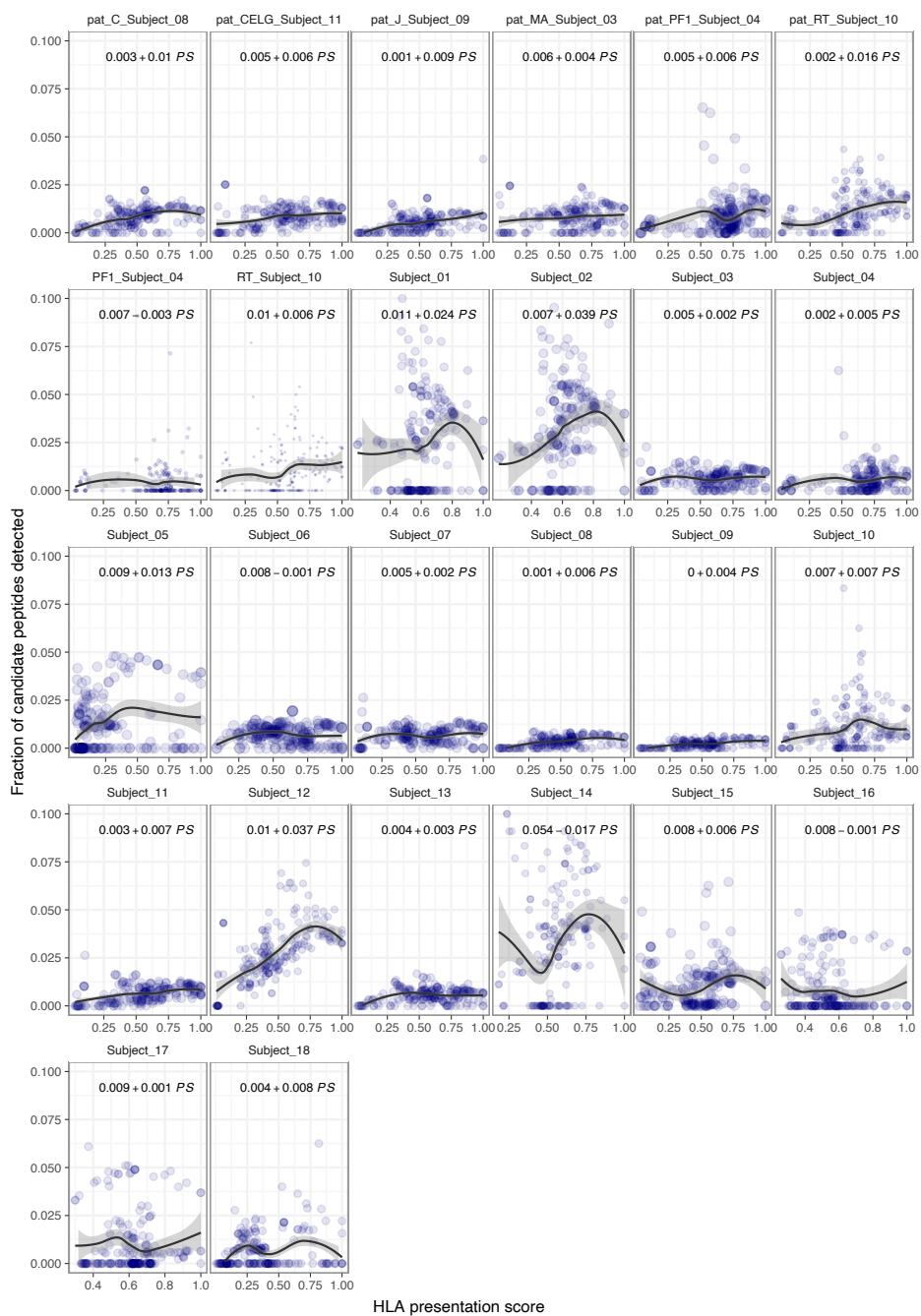
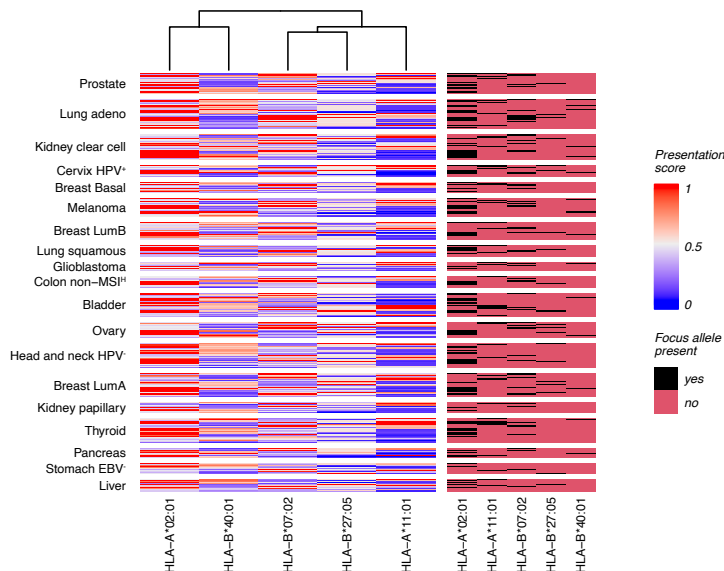


Figure S4.4: Mass-spec based validation of the presentation score for HLA-presentation capability.

Each panel represents a single sample from the Pearson et al. (2016) study³² that performs analyses of HLA-associated peptides by mass spectrometry. Each dot represents a focus allele. Horizontal axis: presentation score for this particular sample for a given focus allele, vertical axis: the fraction of expected peptides under 100% sensitive mass spec technology (i.e., with each peptide from every available protein detectable) that is detected. In general, HLA alleles that show a higher overlap with the predicted binding capacity of the HLA alleles that are expressed by a sample (i.e., a high presentation score) show higher fractions of detected peptides.

A



B

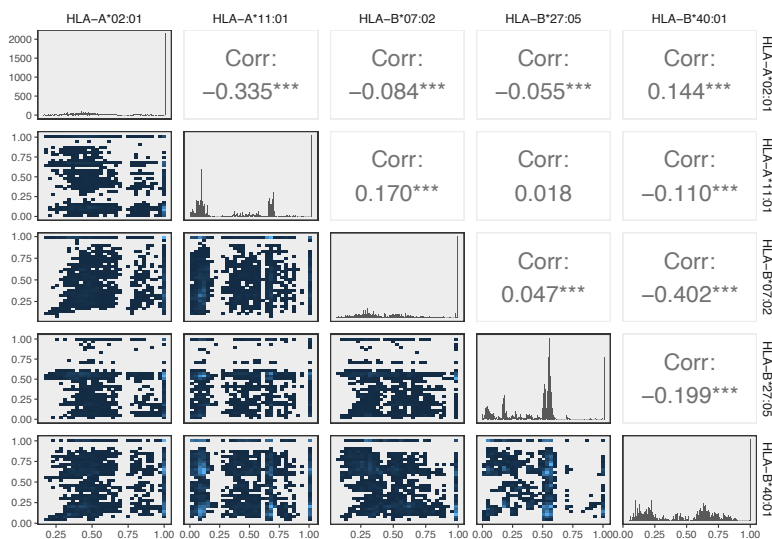


Figure S4.5: Presentation scores are highly variable across focus alleles.

A Heatmap of h across all included samples, computed across all 6 HLA class I alleles in a samples, not leaving out HLA alleles that are genomically lost. Left panel: raw presentation scores. Right panel: binary matrix indicating whether the focus allele is present in the set of a patient's six HLA class I alleles.

B Pairwise correlation matrices of presentation scores. Diagonal elements (histograms) reflect univariate distributions. Sub-diagonal and supra-diagonal elements show a near-absence of correlation between focus alleles.

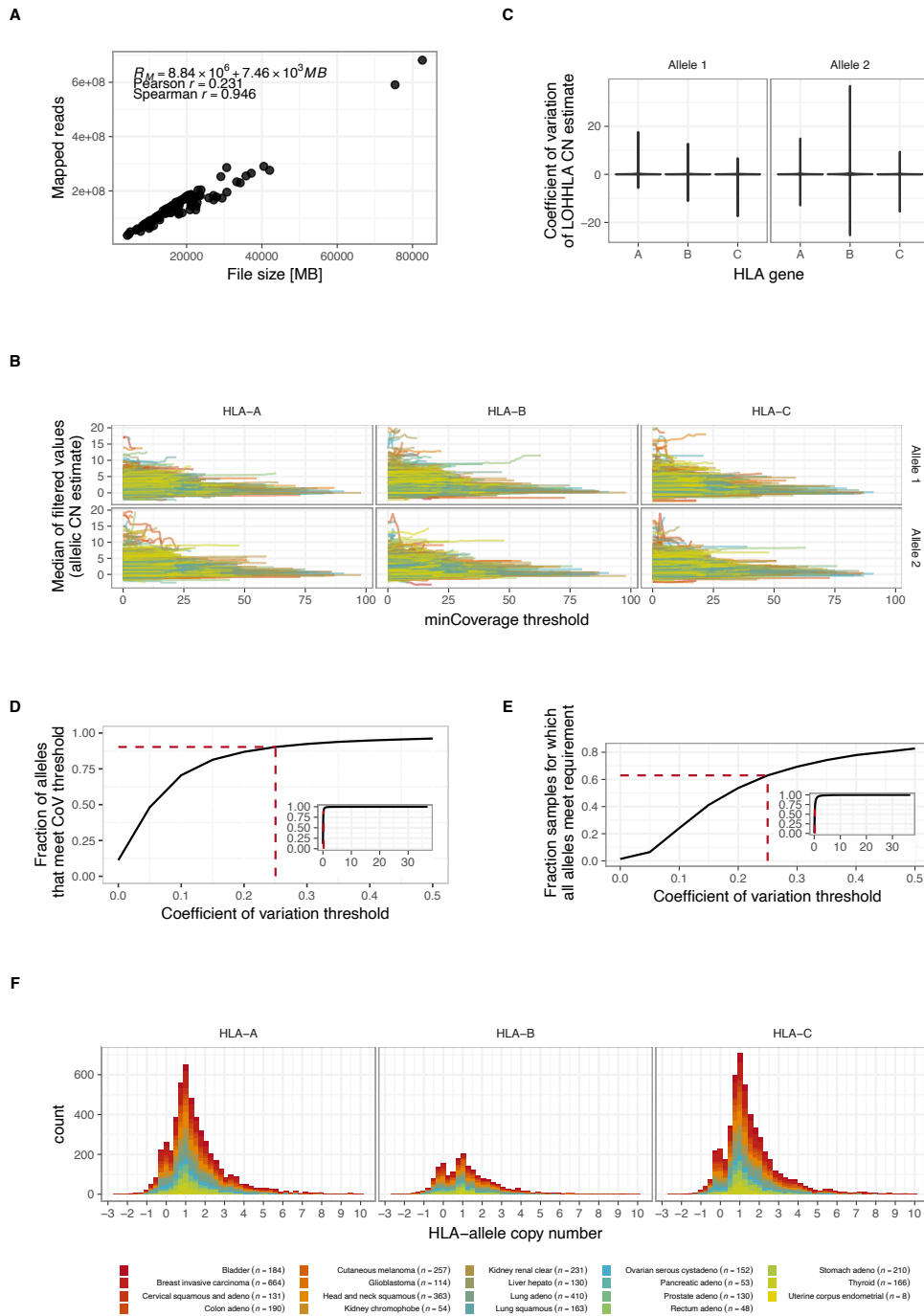


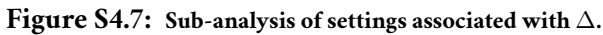
Figure S4.6: Efficient running of LOHHLA for large sample numbers.

A Relationship between the disk size of a TCGA BAM file and its read count content. The strong relationship here allows one to estimate the read content of BAM files from their file size, which, unlike the read count content, can be queried from the TCGA GDC API. Total read count is a required input variable for LOHHLA to run.

B Illustration of the robustness of HLA allele copy number estimates with regards to the LOHHLA minCoverage threshold that defines the minimal number of overlapping reads for a single genomic position to be considered robust enough for use in HLA allele copy number estimates. Each line corresponds to a single copy number estimate.

C Quantification of the robustness in B, using the coefficient of variation (CoV), visualized with violin plots. Most alleles have near-zero estimates, indicating robustness.

D The fraction of HLA allele copy number estimates (6 per patient) with CoV-estimates below the threshold



A As in Figure 4.4A, but after sorting of sub-analyses on Δ rather than q . Sub-analyses with the RNA-expression filter set to 'VE=0' (i.e., requiring $> 0 / 10^6$ total reads to overlap the variant in order to be called expressed) appear enriched among the sub-analyses with the most negative Δ (consistent with neoantigen depletion).

B As in Figure 4.4D, but varying the RNA expression filtering setting on the horizontal axis. Variant-level filtering (VE=0) somewhat strongly affects the mean neoantigen yield rate (r) and also affects Δ . However, it does not affect q at all, indicating the effect on Δ is likely to be artefactual, see supplementary note 1.

C Evidence that stringent neoantigen filtering, resulting in low mean r across all patients, in combination with inhomogeneous distribution of h , causes Δ to be biased. Please see Supplemental Note 1 for associated reasoning. P-values in the lower right corners reflect the linear regression slopes of the $\Delta_{TM B}$ vs. Δ_r regression line.

D As in Figure 4.4B, but for settings associated with the manner in which the neoantigen depletion analysis is performed. Color scale as in A.

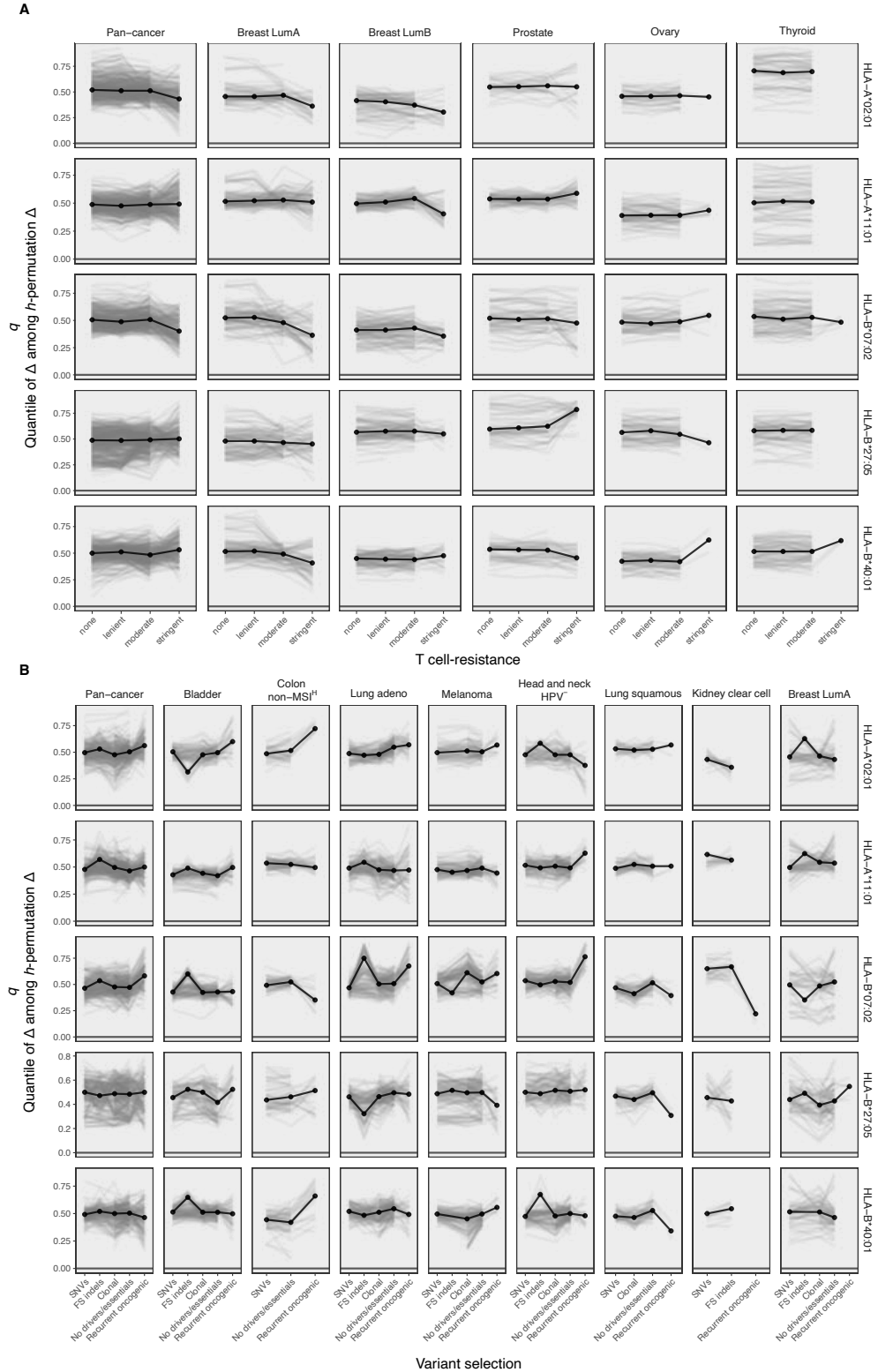


Figure S4.8: Trends of neoantigen depletion are not reproducible across focus alleles.

A As in Figure 4.4E, but split by focus allele and for all tumor types for which stringent filtering of T cell-resistant tumor samples was possible.

B As in Figure 4.4F, but split by focus allele and for all tumor types for which the recurrent oncogenic variant selection was possible.

Supplemental Note

Testing for neoantigen depletion settings that were associated with Δ (rather than q , as used in the main text), we found that stringent RNA expression filtering on the variant level on average led to a negative Δ for a majority of tumor types (Figure S4.7A). Selecting and plotting groupings of identically parameterized sub-analyses, save for the RNA expression filtering setting, confirmed the observation of an effect on Δ and not q (Figure S4.7B).

Interpreting this result, we imagined that stringent neoantigen filtering might enrich for negative Δ but not low q through lowering the mean neoantigen yield rate (r). As low-TMB patients will be more likely to be predicted to have 0 neoantigens, appearing fully depleted of neoantigens, severe reduction of r could render the overall analysis sensitive to imbalances in the distribution of patients over the presentation score (h). This is because any range of h that is relatively sparse in patients will more likely appear to be fully depleted of neoantigens. When relatively few patients populate the lower range of h ($h \in [0, \sim .25]$), Δ could then be biased upwards (i.e., indicating nonsensical enrichment of neoantigens with enhanced neoantigen presentation). Similarly, Δ will be biased downwards with relative patient sparsity at the upper range of h ($h \in [\sim .75, 1]$).

To test this hypothesis, we first computed the mean relative difference in TMB between $h = 1$ and $h = 0$ patients (Δ_{TMB}), much like we did in our primary regression of h against r (Δ), to directly quantify the degree with which h and TMB are correlated. This demonstrated that Δ_{TMB} is strongly associated with Δ (Figure S4.7B), especially for sub-analyses for which the mean r is low, indicating that imbalance in the distribution of h at least partially explains Δ (Figure S4.7C). As h -permutation does not modify the overall h -distribution, permutation Δ s will retain whatever degree of bias that is already present in the original data, such that q is the more robust and informative of the two statistics to assess.

We conclude that, through lowering r , in combination with inhomogeneous distribution of h , variant level expression filtering likely lowered Δ but not q in a manner that is independent from biology.

Supplemental Tables

For the supplemental tables, please see here: [10.1101/2023.06.21.544805](https://doi.org/10.1101/2023.06.21.544805)

DISTINCT SPATIOTEMPORAL DYNAMICS OF CD8⁺ T CELL DERIVED
CYTOKINES IN THE TUMOR MICROENVIRONMENT

Mirjam E. Hoekstra[#], Maarten Slagter[#], Jos Urbanus, Mireille Toebes, Nadine Slingerland, Iris de Rink, Roelof J. C. Kluin, Marja Nieuwland, Ron Kerkhoven, Lodewyk F. A. Wessels and Ton N. Schumacher

[#] These authors contributed equally.

Cancer Cell, 2024. DOI: 10.1016/j.ccell.2023.12.010

Abstract

Cells in the tumor microenvironment (TME) influence each other through the secretion and sensing of soluble mediators, such as cytokines and chemokines. While signaling of interferon γ (IFN- γ) and tumor necrosis factor α (TNF- α) is integral to anti-tumor immune responses, our understanding of the spatiotemporal behavior of these cytokines is limited. Here, we describe a single cell transcriptome-based approach to infer which single or combined signals an individual cell has received. We demonstrate that, contrary to expectations, CD8⁺ T cell-derived IFN- γ is the dominant modifier of the TME relative to TNF- α . Furthermore, we demonstrate that cell pools that show abundant IFN- γ sensing are characterized by decreased expression of TGF- β -induced genes, consistent with IFN- γ -mediated TME remodeling. Collectively, these data provide evidence that CD8⁺ T cell-secreted

cytokines should be categorized into local and global tissue modifiers, and describe a broadly applicable approach to dissect cytokine and chemokine modulation of the tumor microenvironment.

Introduction

Tumors are composed of a diversity of interacting cell types, including tumor cells, fibroblasts, endothelial cells, and a variety of immune cell types. A first type of interactions between the cell populations that jointly form the TME is formed by direct cell-cell contacts, and to describe the effects of such cellular interactions, technologies such as PIC-seq, which allows RNA-sequencing (RNA-seq) of physically interacting cell pairs, have been developed¹. Importantly, next to such direct cell – cell interactions, cells also use soluble factors, such as cytokines, chemokines, and growth factors, to influence the state of surrounding tissue cells².

One of the major cytokine-producing cell compartments in tumor tissue is formed by the CD8⁺ cytotoxic T cell pool, and CD8⁺ T cells have been demonstrated to play a central role in both immune checkpoint blockade³ and adoptive T cell therapies⁴⁻⁶. Upon encounter of antigen-expressing target cells, CD8⁺ T cells release lytic granules containing cytotoxic molecules such as perforin and granzymes in the synapse that is formed between the interacting cells. In addition, T-cell receptor (TCR) signaling leads to the secretion of the cytokines interferon γ (IFN- γ), tumor necrosis factor α (TNF- α) and Interleukin 2 (IL-2), which can, on their own or jointly, induce large-scale alterations in the transcriptome of cells that sense these factors. For example, IFN- γ receptor (IFN- γ R) signaling has been demonstrated to result in increased expression of components of the antigen presentation pathway, enhances expression of immune checkpoint molecules, and can promote recruitment of other immune cells through production of chemokines such as CXCL9, 10 and 11^{7,8}. In addition, IFN- γ and TNF- α have been demonstrated to regulate the activation and maturation state of, amongst others, macrophages and dendritic cells⁹⁻¹². Furthermore, both IFN- γ R and TNF- α receptor (TNF- α R) signaling can, in a context-dependent fashion, contribute to tumor cell senescence¹³, apoptosis^{8,14} and ferroptosis¹⁵. Finally, besides their direct effects on tumor cells, both IFN- γ and TNF- α can also be critical for tumor control through their effects on stromal cells in the tumor vasculature¹⁶⁻²⁰.

In spite of the central role of T cell-produced cytokines in the modulation of cell behavior in the TME, our understanding of the spatiotemporal behavior of CD8⁺ T cell derived cytokines is limited. Specifically, it has not been established whether these signaling molecules differ in their capacity to not only influence target cells in the immediate vicinity of sites of antigen recognition, but also modulate the behavior of cells in the tumor tissue in a more global manner. Earlier work has demonstrated that cytokines may either be secreted in a multidirectional fashion, or can selectively be released in the immune synapse, analogous to the focused release of lytic granules²¹. Specifically, following TCR triggering, membrane-bound TNF- α has been shown to be distributed equally over

the cell membrane, as demonstrated by live imaging of TNF- α on activated murine CD4⁺ T cells²¹. In contrast, in clusters of T cells and antigen-presenting cells, IFN- γ , IL-2, IL-4, and IL-5 were all shown to be localized at the microtubule organizing center (MTOC), consistent with directional release^{21-hyperlinks2323}. Based on this postulated difference in mode of secretion, a more profound effect of TNF- α relative to, for instance, IFN- γ or IL-2 on cells that are distant from the site of antigen recognition could be expected. However, as the size of the cell field in which productive cytokine sensing can occur is also influenced by other parameters, such as cytokine half-life, receptor-mediated clearance, and binding to extracellular components, it has been difficult to predict the extent of long range sensing of different cytokines in the TME²⁴.

Evidence that cytokines can reach (remote) bystander cells that cannot be recognized by T cells directly has been obtained in a number of studies in both viral infection and tumor models²⁵⁻³⁰. Specifically, T cell-secreted IFN- γ in skin and lymph nodes was shown to induce expression of IFN- γ -responsive genes in large regions outside the parasite or virus infected areas^{26,27,31}. Likewise, secretion of IFN- γ and TNF- α by CD4⁺ T cells in tumors has been demonstrated to induce senescence in tumor cells that cannot be directly recognized by T cells¹³. In case of CD8⁺ T cells, long-range sensing of IFN- γ has been observed by, amongst other, intravital imaging of fluorescent IFN- γ R-signaling reporters in mosaic tumors that contain both antigen-positive and antigen-negative tumor areas. Using such fluorescent reporter systems, it was demonstrated that a large fraction of bystander cells senses IFN- γ upon intratumoral T-cell activation, and that IFN- γ sensing can occur in tumor cells at distances over hundreds of micrometers from the site of T cell activation^{29,30}. Collectively, these data on individual cytokines provide an incentive to develop technology to measure and deconvolute the joint effects of multiple cytokines on the TME. In the present study, we set out to generate a strategy that allows the analysis of the effects of a broad set of cytokines simultaneously, and also provides information on the timing of such cytokine exposure. The data obtained demonstrate that, contrary to what would be predicted based on their mode of secretion, IFN- γ is the dominant T cell-secreted modifier of the TME.

Results

In order to measure cytokine sensing in the TME in a manner that is independent of genetic reporter systems, we explored whether gene expression signatures can reliably inform on the type and duration of cytokine exposure. Towards this goal, we exposed human ovarian carcinoma (OVCAR5) cells to different cytokines or cytokine combinations for 2-24 hours and analyzed transcriptomes by bulk RNA-seq (RNA-seq). In line with expectations, this revealed large groups of genes that were selectively induced by IFN- γ (such as *HLA-DRA* and *IRF1*), or TNF- α (such as *CCL20* and *MMP9*). In addition, a gene set was identified that was either preferentially or exclusively induced by the combination of these cytokines (e.g., *UBD* and *CXCL9*) (Figure 5.1a). Furthermore, the relative expression of individual genes in these gene sets provided rich information on cytokine exposure time, distin-

guishing genes with a 'burst-like' expression pattern (e.g., *CCL20* after TNF- α exposure), and genes for which expression showed an exponential increase over time (e.g., *MMP9* after TNF- α exposure) (Figure 5.1a). While gene expression data were rich in terms of the nature and duration of cytokine exposure (Figure 5.1a-b), no substantial differences in gene expression were observed as a function of cytokine concentration beyond a limited concentration range (~10-fold range). Importantly, cell culture medium derived from T cell-tumor cell co-cultures contained high levels of IFN- γ (>100 ng/ml) and TNF- α (\pm 1 ng/ml) and induced a gene expression profile that was highly similar to that observed upon dual IFN- γ *plus* TNF- α exposure (Figure S5.1).

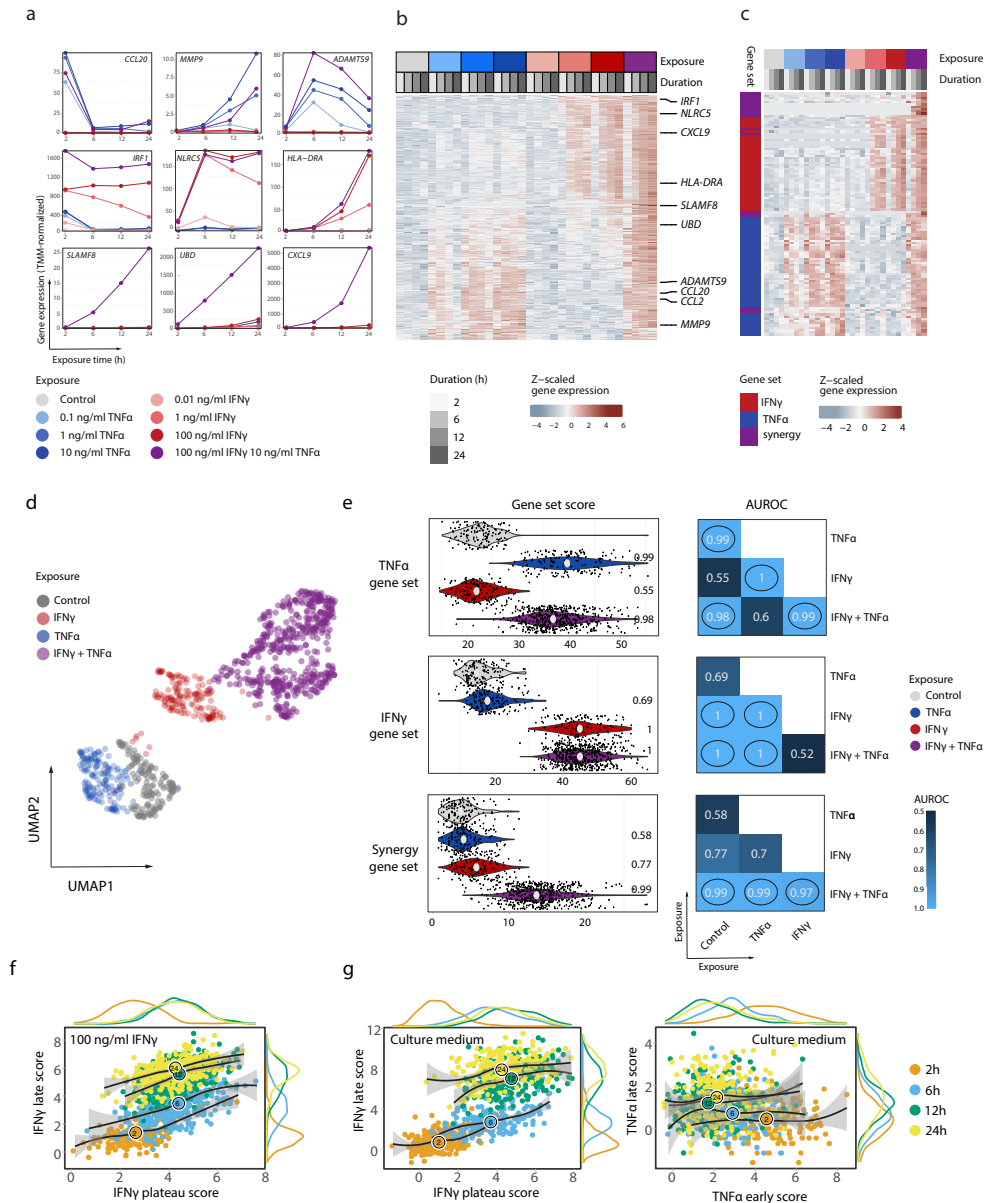


Figure 5.1: Gene expression reports on cytokine exposure

a mRNA expression profiles of selected genes in OVCAR5 cells exposed to indicated concentrations of IFN- γ , TNF- α or their combination, for the indicated duration. Top, middle, and bottom panels depict genes that are primarily responsive to TNF- α , IFN- γ , or TNF- α plus IFN- γ , respectively.

b Heatmap of bulk gene expression values inferred from OVCAR5 cells exposed to indicated concentrations of TNF- α , IFN- γ or TNF- α plus IFN- γ , for indicated durations. Unsupervised hierarchical clustering of data (shown are the 612 genes from the 'cytokine-responsive class', see Figure S5.2b), groups samples by exposure type and then by exposure duration. **c** Heatmap of bulk gene-expression values for mono-responsive genes and synergy genes inferred from *in vitro* stimulated OVCAR5 cells, as in a and b. Unsupervised hierarchical clustering of gene expression data shows a nearly full agreement with assigned gene classes (cluster purity of 0.86).

d UMAP of single cell RNA-seq data of OVCAR5 cells stimulated with indicated recombinant cytokines for 24h.

e Gene set scores for single cell RNA-seq data of *in vitro* cytokine stimulated cells as in a. Left panels: dots represent gene set scores of individual cells; violins represent densities of score distributions. Area under the receiver operator curve (AUROC) values, quantifying how well experimental conditions can be distinguished from the control condition, are depicted. Right panels: Heatmaps showing pairwise distinguishability of indicated experimental conditions (axes) using gene set scores, as quantified using AUROC values. Comparisons for which indicated gene sets are designed to show separation are encircled.

f IFN- γ plateau versus IFN- γ late gene set scores (see Methods) for OVCAR5 cells stimulated with IFN- γ (100ng/ml) for the indicated times. Black lines are LOESS-smoothed curves representing local averages, one per stimulus duration. The ratio of each of the two gene set scores informs on duration of cytokine exposure.

g IFN- γ plateau versus IFN- γ late gene set scores and TNF- α early versus TNF- α late gene set scores for OVCAR5 cells stimulated with culture medium obtained from T cell -tumor cell co-cultures (as in Figure S5.1) for the indicated times as in **F**.

To be able to assign duration and type of cytokine exposure to individual cells, we compiled a set of cytokine informative genes through a combination of model training and manual curation. In brief, to efficiently expand a seed set of ~80 informative genes identified by manual selection, all genes were annotated with descriptive features (see methods), developed to discern cytokine-responsive from unresponsive genes (Figure S5.2a). We next iteratively expanded from this initial seed set by training a machine learning model with the objective of predicting gene classes (e.g., ‘cytokine-responsive’ and ‘cytokine-unresponsive’) for all yet unclassified genes (Figure S5.2b-c). After 10 iterations of model training, gene class prediction, and correction of predicted classes, a set of 612 cytokine-responsive genes was obtained. (Figure S5.2d). 90 of these could be classified as mono-responsive to either IFN- γ ($n=40$) or TNF- α ($n=50$), i.e., with only a single cytokine eliciting a response and IFN- γ *plus* TNF- α eliciting a response that did not substantially deviate from the response to the main excitatory cytokine. This property renders these genes especially useful when simultaneously assessing the spreading behavior of both cytokines (Figure S5.2e and Figure 5.1c) and allows for the summing of expression values for component genes to infer cytokine stimulus (see below). Importantly, comparison of these gene sets to the TNF- α and IFN- γ Hallmark gene sets³² that are frequently used to evaluate signaling revealed only modest overlap (Figure S5.3a). Whereas the newly developed gene sets consisted solely of genes that responded strongly and specifically to the cytokine they were assigned to in OVCAR5 cells, the Hallmark gene sets showed considerable overlap. Furthermore, a sizeable number of TNF- α hallmark genes was shown to respond to IFN- γ and vice versa, creating the potential for incorrect signal inference (Figure S5.3b-d). Next to the set of mono-responsive genes, a set of synergy genes, which are selectively expressed in the presence of both IFN- γ and TNF- α , was identified, providing an independent means to measure co-occurrence of IFN- γ R and TNF- α R signaling (Figure 5.1c). Finally, analysis of gene expression dynamics demonstrated that this approach can inform on cytokine exposure duration (Figure S5.3e).

Having established a number of cytokine- and time-informative gene sets, we tested whether gene expression upon cytokine exposure is informative in single cell (sc) transcriptome data. In an unsupervised analysis, cells exposed to activating concentrations of either IFN- γ , TNF- α , or their com-

bination formed separated clusters, both from control cells and each other, (100% rejection rate on kBET-test³³ with stimuli as batches; median silhouette width: 0.17) (Figure 5.1d). To quantify signal strength for each exposure-specific gene set, we subsequently calculated cell expression scores for all genes that were included in either gene set. Using this strategy on single cell RNA-seq data from cells exposed to activating concentrations of IFN- γ or IFN α revealed a near perfect separation of IFN- γ and TNF- α exposed cells from control cells, as well as a clear separation between cells exposed to the two different stimuli (Figure 5.1e). In addition, exposure to the combination of IFN- γ *plus* TNF- α could be identified with high precision, both by analysis of the separate IFN- γ and TNF- α gene set scores, and by use of the IFN- γ *plus* TNF- α synergy gene set (Figure 5.1e). Furthermore, ability to correctly assign cytokine stimuli was not affected by the experimental strategy (protease digestion, flow cytometric sorting) required to obtain single cell information from tumor material (Figure S5.3f). Finally, use of time-informative gene sets on *in vitro* cultured cells exposed to recombinant cytokines or culture medium from T cell-tumor cell co-cultures at different time points demonstrated the ability of this technique to also infer stimulus duration from single cell data (Figure 5.1f-g and Table S5.6).

Having established methodology for TNF- α and IFN- γ exposure inference in single cells, we subsequently set out to measure the degree of T cell-secreted cytokine sensing by tumor cells in the TME *in vivo*. To this purpose, OVCAR5 tumors that were composed of a large fraction of antigen negative ('bystander') tumor cells that could serve as cytokine sensing reporter cells, plus a small fraction of tumor cells that form targets for neoantigen-specific T cells³⁰ were established in NSG- $\beta 2m^{-/-}$ mice (Figure 5.2a). Following treatment of mice bearing such mosaic tumors with TCR-transduced CDK4_{R>L} neoantigen-specific CD8⁺ T cells, infiltration of CD8⁺ T cells into tumor tissue is observed and upon target cell recognition, cytokine production is initiated in tumor regions composed of antigen-positive tumor cells³⁰. Note that in this setup, all subsequent analyses of cytokine-specific gene set scores by single cell RNA-seq were restricted to antigen negative bystander cells that cannot be recognized by T cells, and hence weren't influenced by direct cell-to-cell killing of antigen-positive tumor cells.

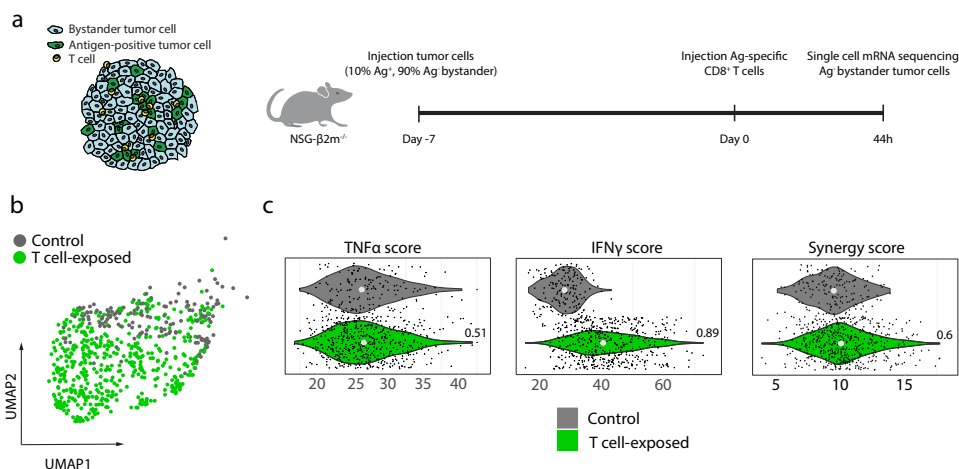


Figure 5.2: Frequent IFN- γ but not TNF- α sensing by bystander tumor cells.

a NSG $\beta 2m^{-/-}$ mice injected subcutaneously with a mixture of 10% CDK4_{R>L} antigen expressing and 90% bystander OVCAR5 tumor cells were treated with either PBS (control) or CDK4_{R>L}-specific CD8⁺ T cells after tumor establishment. Tumors were harvested 44h after treatment, and bystander tumor cells were analyzed by single cell RNA-seq.

b UMAP of single cells based on gene expression in the “cytokine-responsive” gene class, as described in Figure S5.2b.

c TNF- α , IFN- γ , and synergy gene set scores of single cells derived from OVCAR5 tumors. Dots represent gene set scores of individual cells, violins represent densities of score distributions. Numeric values reflect AUROC values that quantify separability between experimental conditions. Note that IFN- γ , but not TNF- α , gene set scores are increased in the T cell-exposed condition as compared to the control condition.

Having established that T cell infiltration and activity are first detected around 16 hours and increase up to 44 hours after T cell transfer (Figure S5.4a), we analyzed bystander tumor cells by single cell RNA-seq at 16-44 hours after T cell transfer. Unsupervised clustering of bystander cells derived from 44h T cell-exposed tumors and control tumors demonstrated a separation of a large fraction of bystander cells obtained from T cell-exposed tumors (Figure 5.2b). Thus, the presence of a tumor-reactive CD8⁺ T cell compartment substantially modifies the transcriptome of a considerable part of bystander tumor cells in the TME. Importantly, assignment of cells to different cytokine exposure conditions revealed that a large fraction (70.4%) of bystander tumor cells in T cell-treated mice showed a pronounced IFN- γ gene set score, whereas such IFN- γ sensing was largely absent in tumor cells from control mice (5.3% of cells) (Figure 5.2c and Figure S5.4b). In contrast, presence of a tumor-reactive T cell compartment did not measurably increase the fraction of TNF- α sensing cells, with 5.3% of tumor cells classified as TNF- α sensing in both control and T cell-exposed tumors (Figure 5.2c and Figure S5.4b). In addition, tumor cells displaying high IFN- γ gene set scores did not show elevated TNF- α gene set scores (Fig S4b). As a second test of *in vivo* TNF- α exposure, we calculated synergy gene set scores, which independently inform on the sensing of the combination

of IFN- γ *plus* TNF- α (Figures 5.2c and S5.4b). Also by this metric, the presence of a tumor-reactive T-cell compartment did not result in TNF- α sensing by an appreciable fraction of bystander tumor cells. As a majority of TNF- α genes displays a burst-like, early, expression pattern, we next assessed cytokine sensing at the 16h timepoint, at which measurable T cell infiltration is just visible (Figure 5.3a). Already at this time point, a subset of bystander tumor cells derived from T cell-exposed tumors separated from bystander tumor cells in control tumors (Figure 5.3b). However, neither the use of the entire TNF- α mono reporter gene set (Figure 5.3c, left), nor the use of the TNF- α early time-informative gene set, showed an appreciable TNF- α sensing signal (Figure 5.3d, right). Note that T cells derived from such tumors did retain the capacity to produce TNF- α , indicating that the lack of an appreciable tumor cell population that showed TNF- α sensing was not explained by impaired cytokine production (Figure S5.4c). As a side note, application of the time reporting IFN- γ -responsive gene sets demonstrated that tumor cells isolated 44 hours after T-cell infusion that show a given expression of the IFN- γ “plateau gene set score” on average showed a slightly increased expression of the IFN- γ “late gene set score”, as compared to tumor cells analyzed 16 hours after T-cell infusion (Figure 5.3d, left). To test whether bystander tumor cells did retain the capacity to respond to TNF- α *in vivo* when this cytokine is present, we intratumorally injected tumors with recombinant cytokines. Importantly, an evident TNF- α signal was observed upon injection of either TNF- α (with 53.4% of cells surpassing 95th percentile of control) or TNF- α *plus* IFN- γ (51.8%) (Figure 5.3e and Figure S5.4d). Furthermore, a pronounced synergy signal was selectively observed upon intratumoral injection of TNF- α *plus* IFN- γ (49.4% of cells), likewise indicating a high sensitivity to detect TNF- α R signaling *in vivo*.

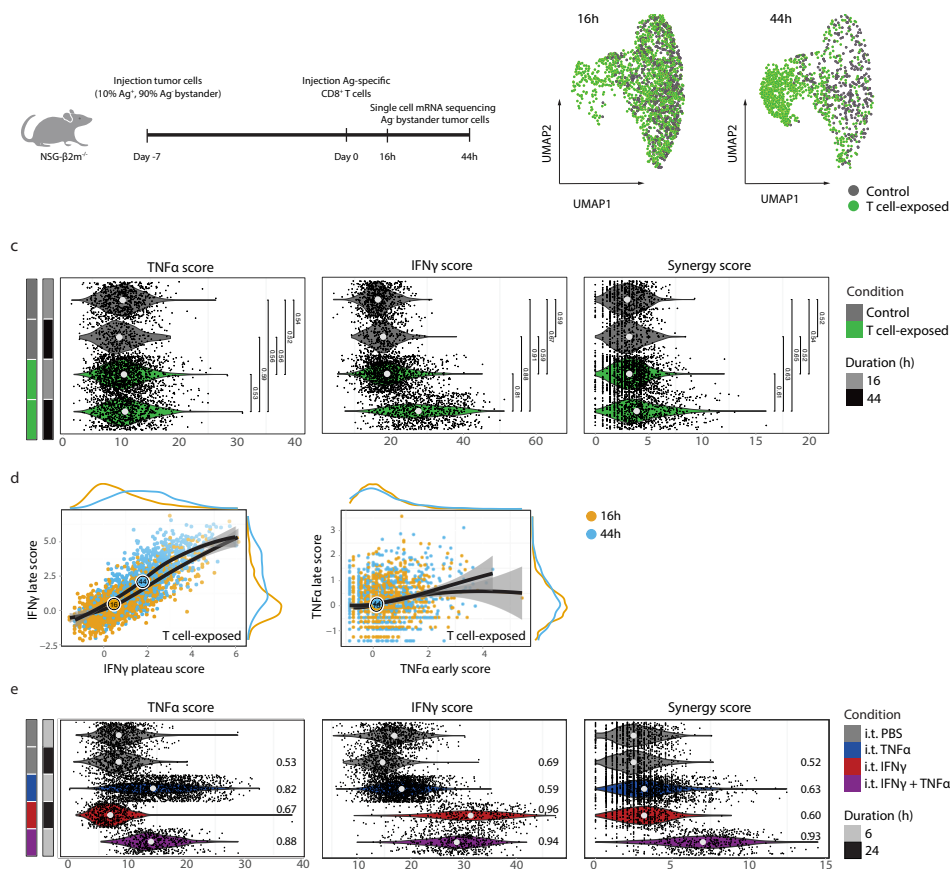


Figure 5.3: Appreciable IFN- γ but not TNF- α sensing by bystander tumor cells early after T-cell activation.

a NSG- $\beta 2m^{-/-}$ mice injected subcutaneously with a mixture of 10% CDK4_{R>L} antigen expressing and 90% bystander OVCAR5 tumor cells were treated with PBS (control) or with CDK4_{R>L}-specific CD8⁺ T cells after tumor establishment. Tumors were harvested 16h or 44h after treatment, and bystander tumor cells were subjected to single cell RNA-seq.

b UMAP of single cell sequencing data from bystander cells of control and T cell-exposed OVCAR5 tumors harvested 16 or 44h after treatment, based on genes in the “cytokine-responsive” class (Figure S5.2b).

c Violin plots of TNF- α , IFN- γ , and synergy gene set scores of cells derived from OVCAR5 tumors, as described in a. Dots represent gene set scores of individual cells, violins represent densities of score distributions. Numeric values reflect AUROC values that quantify separability between experimental conditions.

d Scatter plots of time-informative gene sets for *in vivo* single cell data described in a. To remove gene expression effects due to exposure duration independent from T-cell exposure, depicted gene set scores were normalized to control (PBS treated) counterparts in a duration-matched fashion (Methods).

e TNF- α , IFN- γ , and synergy gene set scores of OVCAR5 tumor cells derived from tumors injected with indicated recombinant cytokines. Cytokine exposure times were chosen based on maximal change in expression of cytokine specific responsive genes after *in vitro* cytokine exposure (Figure 5.1b). Numeric values reflect AUROC values that quantify separability between experimental conditions, as in Figure 5.1e.

To test whether the observed difference in long-range IFN- γ and TNF- α sensing also occurs in syngeneic tumor models, in which not only tumor cells but also infiltrating immune cells can respond to T cell-secreted cytokines^{34,35}, we compiled responsive gene sets from cytokine-stimulated mouse NRAS mutant melanoma (NMM) cells (Figure S5.5a). Application of these gene sets to bystander cells derived from mosaic NMM tumors demonstrated that a large fraction of bystander tumor cells (38.0%) responded to IFN- γ in T cell-treated mice, as compared to bystander tumor cells in mice that did not receive antigen-specific T cells (5.1%), or in mice in which antigen was lacking (5.7%). In contrast, presence of a tumor-reactive CD8⁺ T-cell compartment did not induce sensing of TNF- α by bystander tumor cells (4.5% responding cells vs 5.1% and 5.7% in the two controls, respectively) (Figure 5.4a-b, Figure S5.5b). For NMM tumors, but not for OVCAR5 tumors, it is possible that cytokine-induced cell death results in a slight underestimate of the fraction of cells encountering the combination of IFN- γ and TNF- α signals. However, for both models, the observed bias towards IFN- γ sensing was not predominantly explained by TNF- α induced cell death (Figure S5.5c). Collectively, these data demonstrate in two different mouse models, and using gene sets that either report on the sensing of individual cytokines or on the combination of IFN- γ *plus* TNF- α , that widespread sensing is restricted to T cell-derived IFN- γ .

The ability to identify bystander tumor cells that have sensed IFN- γ *in vivo* makes it possible to test whether such sensing is associated with additional changes in cell state. To explore this, we used Milo³⁶ to identify transcriptionally similar cells (so-called neighborhoods) in the mouse NMM melanoma data. 64 out of 128 neighborhoods were enriched for bystander tumor cells derived from T cell-exposed tumors (hereafter referred to as ‘T cell-exposed neighborhoods’) relative to bystander cells from the PBS control condition (figure 5.4c) jointly comprising 74.3% of bystander tumor cells from T cell-exposed tumors. As a control, none of these neighborhoods were enriched or depleted for bystander tumor cells derived from T cell treated tumors in which antigen was lacking (Fig 4d, top). As expected, T cell-exposed neighborhoods showed a prominent IFN- γ -sensing profile but were also characterized by reduced expression of a second gene set that showed considerable overlap with genes induced by *in vitro* TGF- β stimulation of NMM melanoma cells (Figure 5.4d). Analysis of selected TGF- β responsive genes from bulk MNN RNAseq data (Figure S5.5a bottom panel) showed that a majority of these (47 out of 70) were negatively correlated with T-cell pressure (Figure 5.4e). However, a large fraction of TGF- β induced genes also appeared to show reduced expression upon IFN- γ exposure (Figure S5.5a), making it difficult to unambiguously ascribe this transcriptional response to lowered TGF- β sensing in T cell neighborhoods using solely gene signatures. To disentangle the transcriptional effects of co-occurring cytokines that regulate partly overlapping gene sets, we employed transcriptional deconvolution of cell neighborhoods, aiming to reconstruct their transcriptomes by algorithmically identifying optimal mixing weights of whole-transcriptome bulk RNAseq profiles (Methods), akin to the CIBERSORT approach³⁷. A high similarity was observed between control neighborhoods and profiles of TGF- β -stimulated and unstimulated cells. In contrast, T cell-exposed neighborhoods more strongly resembled IFN- γ expression profiles (Figure 5.4f). Notably, omission of TGF- β -exposed reference profiles from this analysis increased recon-

struction error and predominantly did so for control neighborhoods (Figure 5.4g, left). In contrast, exclusion of IFN- γ -profiles specifically increased reconstruction error of T cell-exposed neighborhoods (Figure 5.4g, right). Collectively, these data demonstrate that T-cell pressure modulates bystander tumor cells towards transcriptional activity that is consistent with IFN- γ -sensing and with reduced TGF- β -sensing.

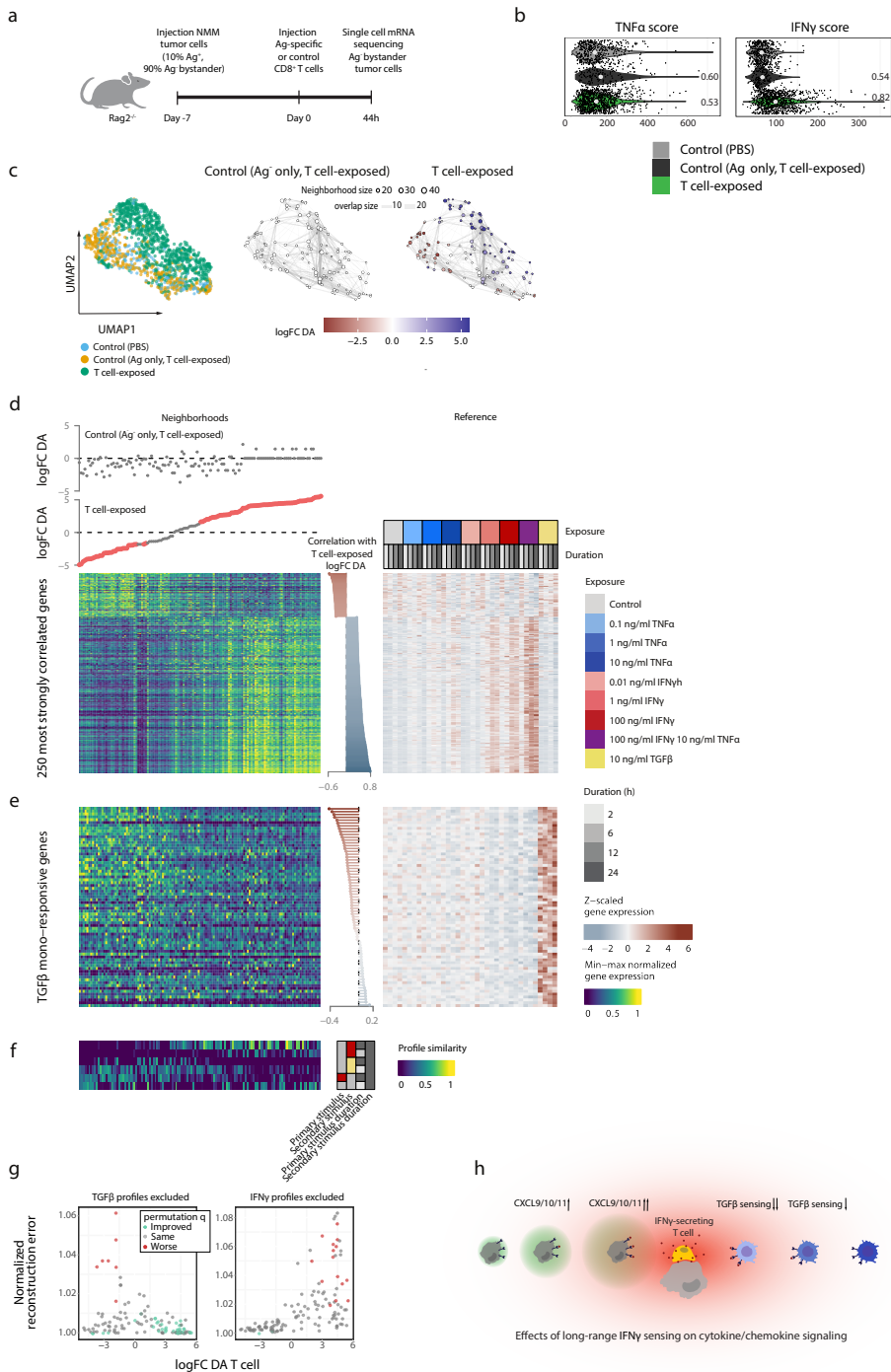


Figure 5.4: Figure 4. Frequent IFN- γ sensing in a syngeneic tumor model and relationship with reduced TGF- β sensing.

a Rag2^{-/-} mice were injected subcutaneously with a mixture of 10% OVA antigen expressing and 90% Ag⁻ bystander NMM tumor cells, or with Ag⁻ NMM tumor cells only, and, following tumor establishment, were

treated with either PBS (control) or OT-1 CD8⁺ T cells, as indicated. Ag⁻ bystander tumor cells were harvested for single cell RNA-seq analysis 44h after treatment.

b TNF- α and IFN- γ gene set scores, determined using the genes shown in **a**, for the T cell-exposed condition (green) and the two control conditions (T cell-exposed -Ag⁻ bystander NMM tumor cells only tumors, and PBS treated tumors, shades of gray). Formatting as in Figure 5.1e.

c Left panel: UMAP of NMM melanoma single cell data, as described in panels **a** & **b**. Middle and right panels: a Milo model³⁶ was fitted to the data to test for enrichment or depletion for any of the experimental conditions in neighborhoods of transcriptionally similar cells. Non-significantly imbalanced neighborhoods (Spatial FDR > 0.05), as well as homogeneous neighborhoods, are colored white.

d Left panel: heatmap of top 250 genes (rows) most strongly correlated (Spearman correlation) with enrichment for the T cell-exposed condition in cell state neighborhoods (columns) of transcriptionally similar cells. Depicted values are neighborhood averages. Neighborhoods are ordered according to compositional enrichment of cells from the T cell-exposed condition. Top panels show log fold change in differential abundance (logFC DA) for the indicated experimental condition relative to control condition. Right panel: heatmap showing bulk RNAseq gene expression profiles of NMM cells exposed to indicated cytokines for the same genes as in the heatmap in the left panel, ordered identically.

e As in **d**, but for TGF- β responsive genes selected on bulk RNAseq data.

f Deconvolution mixing weights of neighborhoods in an independent bulk RNAseq experiment (Figure S5.5a, right). Neighborhoods ordered as in **d**. Only the 6 out of 28 most highly selected reference profiles are shown, jointly comprising 94% of all assigned similarity.

g (Left) Increase in reconstruction error when the 17 reference profiles with TGF- β are omitted as compared to when all 28 profiles are included. Permutation testing was employed to test whether increase in reconstruction error could be explained by a lower number of reference profiles (Methods). (Right) As left, but omitting the 17 reference profiles with IFN- γ .

h Model visualizing secondary effects of long range IFN- γ sensing. In parallel to the mechanism in which long range IFN- γ sensing leads to generation of, for instance, CXCL9/10/11 chemokine fields and subsequent increased immune cell infiltration, long range IFN- γ sensing may result in secondary changes in the TME by decreasing TGF- β sensing.

Discussion

Next to signaling events induced by direct cell–cell contact, tumor cell behavior is modulated through the sensing of soluble mediators, such as chemokines and cytokines, offering possibilities for long range communication. Here, we describe and validate a single cell sequencing-based approach to identify such long-range communication, and also the secondary changes that are associated with it. Key components of this approach are the generation of bespoke gene sets that report on cell exposure to a given cytokine or cytokine combination, and also the employment of multivariate modeling in case signal-specific reporter genes are unavailable. We demonstrate in both humanized and syngeneic tumor models that CD8⁺ T cells predominantly modulate the behavior of the tumor mass through IFN- γ release, while no substantial evidence for widespread TNF- α sensing is obtained. This lack of TNF- α sensing is observed in spite of ongoing IFN- γ sensing by a large fraction of the tumor mass, and hence continuous T-cell activity (note that a single intratumoral application of IFN- γ results in just a transient burst of IFN- γ sensing, Figure S5.4d and S5.5a). In addition, we note that bystander tumor cells do retain the capacity to respond to TNF- α *in vivo* when this signal is artifi-

cially provided (Figure 5.3e). In theory, the methodology that we describe is subject to inferential bias in case the studied signals influence cell survival. In our case, such cell survival effects do not form a significant confounder (Figure S5.5c), but it is important to be aware of this possibility when examining other cell models. The cumulative CD8⁺ T cell-derived IFN- γ and TNF- α levels will vary across tumors, depending on e.g., the fraction of tumor cells presenting relevant antigen and T cell density. Importantly though, we consider it likely that, unless local TME signals would differentially influence production of either cytokine, the ratio between T cell produced TNF- α and IFN- γ will be constant due to their shared dependence on TCR-triggering. Hence, our observation of a differential reach of CD8⁺ T cell-derived IFN- γ and TNF- α is expected to generalize to unseen settings.

Prior work has demonstrated that exposure to IFN- γ and TNF- α can influence tumor control by, for instance, inhibiting growth of antigen loss variants^{13,18,29,30} and modifying behavior of tumor stromal cells^{16,38}. Here we demonstrate that, whereas T cell-derived IFN- γ modulates the behavior of a large fraction of antigen-negative cells in the TME, such global effects are not observed for TNF- α . These observations lead us to propose a distinction between cytokines that act as local versus global modifiers of the TME. Importantly, modification of tumor growth through the global TME modifier IFN- γ may be expected to already occur in settings in which T cell activity is heterogeneous and restricted to smaller areas of the tumor cell mass. In contrast, the effect of local TME modifiers such as TNF- α may be most apparent in case of a stronger and more homogeneous intratumoral T cell response. Notably, should other cell types show a drastically different ratio of IFN- γ and TNF- α production, it would be of interest to evaluate differential sensing in these settings.

Finally, the ability to detect sensing of individual cytokines and chemokines makes it possible to determine whether such sensing is associated with additional alterations in cell state. In the current work, we demonstrate that cell states induced by T-cell activity are not only consistent with abundant IFN- γ sensing but also with decreased TGF- β -induced gene expression. Conceivably, intratumoral IFN- γ sensing could result in reduced availability of bioactive TGF- β , for instance through induction of a more pro-inflammatory macrophage state^{39,40}. In addition, we provide evidence for mutual negative regulation between the two cytokines (Figure 5.4h). These data add to an emerging view on the role of tumor-reactive CD8⁺ T cells, in which TCR signaling induced cytokine secretion, and in particular IFN- γ secretion, results in a global alteration of the tumor microenvironment.

Methods

Tumor cell culture and viral transductions

Human ovarian carcinoma OVCAR5 cells (F. Scheeren, The Netherlands Cancer Institute, The Netherlands) and murine NRAS^{Q61R} mutant melanoma cells (NMM) (Norman Sharpless, University of North Carolina, USA,⁴¹) were cultured at 37°C/ 5% CO₂ in IMDM (Gibco) supplemented with 10%

FCS (Sigma), 100 U/ml penicillin (Roche), 100 μ g/ml streptomycin (Roche), and GlutaMax (Gibco, 1x). Identity of OVCAR5 cells was validated by short tandem repeat analysis, STR data were not available for the NMM cell line. The following vectors were utilized: the CDK4_{R>L}-GFP-pMX, CMV-pMX, and IGS-PCDH vectors as described in ³⁰, the OVA-mPlumb-pLenti vector as described in ⁴², and the CDK4_{R>L} specific TCR (clone 17, NKI12)-pMP71 vector, as described in ⁴³. For retroviral transduction of human cells and mouse cells, FLYRD18 packaging cells (ECACC no. 95091902), and Phoenix-ECO packaging cells (ATCC, CRL-3214), were plated into 6 well plate dishes at 0.5×10^6 cells per well, respectively. After 24h, cells were transfected with 3 μ g of one of the indicated retroviral vectors using X-tremeGENE (Roche), according to the manufacturer's protocol. After 48h, virus supernatant was harvested, filtered through a 0.45- μ m filter and added to tumor cells in the presence of 8 μ g/ml polybrene (Sigma) in a 1:1 dilution in medium. For lentiviral transductions, HEK293T cells (ATCC, CRL-3216) were plated at 3×10^6 cells per 10 cm dish. After 24h, cells were transfected with 8 μ g of one of the above indicated lentiviral plasmids, plus the lentiviral packaging and envelope plasmids psPAX (Addgene #12260) and pMD2.G (Addgene #12259) (3 μ g each) using X-tremeGENE (Roche), according to the manufacturer's protocol. 2-3 days after transfection, supernatant of transfected cells was harvested, filtered through 0.45- μ m filters, and added to OVCAR5 or NMM cells in a 1:1 dilution in medium. Antigen-positive GFP⁺ OVCAR5 cells were generated by retroviral transduction with the pMX-CDK4_{R>L}-GFP vector. Antigen-positive mPlumb⁺ NMM cells were generated by lentiviral transduction with the pLenti-OVA-mPlumb vector. Antigen-negative CFP⁺ bystander OVCAR5 and NMM cells were generated by retroviral transduction with the pMX-CFP vector. After transduction, indicated cell populations were sorted on a FACSaria Fusion (BD biosciences) to >90% purity. Ag⁻CFP⁺IGS reporter cells were generated as described ³⁰.

Generation and culture of TCR-modified T cells

Retroviral transduction and culture of human T cells was performed as described previously ³⁰. To obtain murine GFP⁺ OT-1 CD8⁺ T cells, spleens from C57BL/6;UBC-GFP;OT-I mice were passed through 70 μ m strainers (Falcon) to obtain single cell suspensions. Splenocytes were then negatively enriched with the Mouse CD8 T Lymphocyte Enrichment Set (BD Biosciences) and activated at 1×10^6 cells per 24 well for 48h with 2 μ g/ml Concanavalin A (Merck) in RPMI 1640 supplemented with 8% FCS, penicillin/streptomycin, 50 μ M β -mercapto-ethanol (Gibco), 10 ng/ml IL-2 (Immunotools), 0.5 ng/ml IL-7 (Immunotools) and 1 ng/ml IL-15 (Immunotools). After 48h, cells were spun down and taken up in fresh medium at a concentration 1×10^6 cells/ml. Cells were kept at a concentration of 1×10^6 /ml, with refreshment of media every 24h, for 2-5 days before adoptive transfer.

In vitro cytokine stimulation

Ag⁻CFP⁺ tumor cells were plated at 200,000 cells per well in 6-well plates for 24h and were then treated with either human IFN- γ (Invitrogen), human TNF- α (Peprotech), murine IFN- γ (Thermo

Fisher), murine TNF- α (Peprotech) or murine TGF- β (ebioscience) the indicated combination, or were treated with culture medium from T cell -tumor cell co-cultures (see below), at the indicated concentrations or dilutions. At the indicated times, cells were harvested and used for bulk RNA-seq or single cell RNA-seq, as indicated. For bulk RNA-seq, cells were lysed in RLT lysis buffer (Qiagen) and stored at -80°C before sequencing. The bulk mRNA data sets obtained from these *in vitro* cytokine-stimulated cells at different time points are referred to as the “OVCAR5 bulk RNA-seq reference data set” and “NMM bulk RNA-seq reference data set” throughout the manuscript. For single cell RNA-seq, cells were stained with TotalSeq Hashtag antibodies (TotalSeq-B, Biolegend) and pooled, using an equal number of cells from each sample, to form one pool of cells for single cell RNA-seq analysis. Cell death and total cell counts were analyzed at 16h or 44h after treatment by IR-Dye staining and subsequent flow cytometry using AccuCountBlank 15.2- μ m beads (Spherotech). To obtain culture medium from T cell -tumor cell co-cultures, Ag⁺ GFP⁺ OVCAR5 tumor cells were plated at 2×10^6 per 10cm culture dish. After 1 day, 4×10^6 CDK4_{R>L}-specific CD8⁺ T cells were added, and culture medium was harvested after 24h, filtered through a 0.45 μ m filter (GE) and stored at -80°C.

Mice

NOD-scid Il2 γ ^{null} β 2m^{null} (NSG- β 2m^{-/-}), C57BL/6;RAG2 KO, C57BL/6;UBC-GFP, and C57BL/6;OT-I mice were obtained from Jackson Laboratories. UBC-GFP and OT-I mice were crossed to obtain GFP-OT-I donor mice for adoptive cell transfer experiments. All animal experiments were approved by the Animal Welfare Committee of The Netherlands Cancer Institute (NKI), in accordance with national guidelines. All animals were maintained in the animal department of NKI, housed in individually ventilated cage (IVC) systems under specific pathogen-free conditions and received food and water ad libitum. Mice were used at 8 to 26 weeks of age.

In vivo tumor models

8×10^6 OVCAR5 cells or 2×10^5 NMM cells were injected subcutaneously into the flank of NSG- β 2m^{-/-} mice or C57BL/6;RAG2^{-/-} mice, respectively, in 50 μ l PBS (Gibco) and 50 μ l matrigel (Corning), using the indicated mixtures of tumor cell variants. At day 7-8 after tumor inoculation, tumor-bearing mice received an intravenous injection of either 100 μ l PBS or 5×10^6 CDK4_{R>L} TCR-transduced CD8⁺ T cells or GFP⁺ OT-1 CD8⁺ T cells in 100 μ l PBS, as indicated. On days 0, 1 and 2 after T cell transfer, NSG- β 2m^{-/-} mice received injections of 7.2×10^5 IU IL-2 dissolved in 200 μ l PBS, twice daily, with an interval of 6–12 h between injections, to support T-cell engraftment. At the indicated times after T-cell transfer, mice were sacrificed and tumors were harvested. Harvested tumors were manually minced and enzymatically digested in RPMI medium (Gibco) supplemented with 5 Wünsch units/ml TH Liberase/ml (Roche) 25 μ g/ml DNase I (Roche) for 20 min at 37 °C under continuous shaking. Subsequently, cell digests were filtered through a 70- μ m strainer (Falcon)

and single-cell suspensions were stained with IR-Dye (Invitrogen) and TotalSeq Hastag antibodies (TotalSeq-A or B, Biolegend). Cells from each sample were combined at equal numbers, and CFP⁺ (i.e., bystander) tumor cells were sorted from this cell pool on a FACSaria Fusion (BD biosciences) and analyzed by single cell RNA-seq (see below). To measure cytokine production of intratumoral T cells *ex-vivo*, tumor cells from digested tumors were subsequently cultured *in vitro* in the presence of Ag⁻ CFP⁺ tumor cells (for digests from control tumors) or CDK4_{R>L} Ag⁺ tumor cells (for digests from Ag⁺/Ag⁻ mixed tumors) for 3 hours in the presence of Golgi-plug (BD biosciences) to block cytokine secretion. Cells were subsequently stained for intracellular IFN- γ and TNF- α and analyzed by flow cytometry.

Intratumoral cytokine injections

Where indicated, tumors of $\geq 150 \text{ mm}^3$ size were intratumorally injected with 15ul PBS containing the indicated cytokines (100 ng IFN- γ , 10 ng TNF- α or 100 ng IFN- γ *plus* 10 ng TNF- α per ml of tumor mass), using a Veo™ insulin syringe with a BD Ultra-Fine™ 6mm x 31G needle (BD biosciences). At the indicated times after injection, tumors were harvested, digested, sorted, and bystander tumor cells were analyzed by single cell RNA-seq as described below.

Flow cytometry

For analysis of immune infiltrates in NSG- $\beta 2m^{-/-}$ mice, cells were stained with fluorochrome-labeled anti-human CD3 antibody (clone OKT3; BD biosciences) in FACS buffer (PBS supplemented with 0.5% w/v bovine serum albumin (Sigma) and EDTA (2 mM, Life Technologies)) for 20–30 min at 4 °C, while protected from light. For analysis of cytokine secretion of intratumoral T cells, cells were stained for anti-human CD3 (clone OKT3; BD biosciences), anti-mouse TCR β constant domain (clone H57-597; BD Biosciences), anti-human IFN- γ (clone 4S.B3; BD Biosciences) and anti-human TNF- α (clone MAb11, eBioscience). After incubation, cells were washed twice with FACS buffer before resuspension in FACS buffer for analysis. IR-Dye (Invitrogen) was used to allow for live cell selection.

Bulk RNA-seq

Total RNA was isolated using the RNeasy Mini Kit (74106, Qiagen), including an on-column DNase digestion (79254, Qiagen), according to the manufacturer's instructions. RNA quality and quantity was assessed on the 2100 Bioanalyzer instrument, following the manufacturer's instructions "Agilent RNA 6000 Nano" (G2938-90034, Agilent Technologies). Total RNA samples having RIN values > 8 were subjected to TruSeq stranded mRNA library preparation, according to the manufacturer's instructions (Document # 1000000040498, Illumina). Stranded mRNA libraries were ana-

lyzed on a 2100 Bioanalyzer instrument, following the manufacturer's protocol "Agilent DNA 7500 kit" (G2938-90024, Agilent Technologies), diluted to 10nM and pooled equimolar into multiplex sequencing pools for sequencing on HiSeq 2500 and NovaSeq 6000 instruments (Illumina). HiSeq 2500 single-end sequencing was performed using 65 cycles for Read 1, and 10 cycles for Read i7, using HiSeq SR Cluster Kit v4 cBot (GD-401-4001, Illumina) and HiSeq SBS Kit V4 50 cycle kit (FC-401-4002, Illumina). NovaSeq 6000 paired-end sequencing was performed using 54 cycles for Read 1, 19 cycles for Read i7, 10 cycles for Read i5, and 54 cycles for Read 2, using the NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles) (20028401, Illumina).

Single cell gene expression library generation and sequencing

Single cell suspensions were diluted to a final concentration of 1,000 cells/ μ l in 1xPBS containing 0.04% weight/volume BSA. The Chromium Controller platform of 10X Genomics was used for single cell partitioning and barcoding. Per single cell suspension, each cell's transcriptome was bar-coded during reverse transcription, pooled cDNA was amplified and Single Cell 3' Gene Expression libraries and Cell Hashing libraries via Feature barcode technology were prepared, according to the manufacturer's protocol (CG000183, CG000206 and CG000317, 10X Genomics). All libraries were quantified on a 2100 Bioanalyzer Instrument following the Agilent Technologies Protocol (Agilent DNA 7500 kit, G2938-90024). Sequence library pools were composed and quantified by qPCR, according to the KAPA Library Quantification Kit Illumina® Platforms protocol (KR0405, KAPA Biosystems). HiSeq 2500, NextSeq 550 or NovaSeq 6000 Illumina sequencing systems were used for paired-end sequencing of the Single Cell 3' Gene Expression libraries and Cell Hashing libraries, respectively, at a sequencing depth of between 20,000-60,000 reads /cell and approximately 3,500 reads/cell. HiSeq 2500 paired-end sequencing was performed using 100 cycles for Read 1, 8 cycles for Read i7, and 100 cycles for Read 2, using HiSeq PE Cluster Kit V4 (PE-401-4001, Illumina) and multiple HiSeq SBS Kit V4 50 cycle kits (FC-401-4002, Illumina). NextSeq 550 paired-end sequencing was performed using 28 cycles for Read 1, 10 cycles for Read i7, and 54 cycles for Read 2, using the NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (20024906, Illumina) and NextSeq 500/550 Mid Output Kit v2.5 (150 Cycles) (20024904, Illumina). Novaseq 6000 paired-end sequencing was performed using 28 cycles for Read 1, 10 cycles for each Read i7 and i5, and 90 cycles for Read 2, using the NovaSeq 6000 S2 Reagent Kit v1.5 (100 cycles) (20028316, Illumina). Gene expression and antibody sequencing reads were mapped to the GRCh38 human reference genome (refdata-cellranger-GRCh38-3.0.0) and antibody reference sequences, respectively, using CellRanger Version 5.0.1 in multi mode (10x Genomics) with default parameters. The genomic sequence of the Katushka fluorescent protein (named as ENSG0000055555) was added to the human reference prior to mapping.

Bulk RNA-seq data preprocessing

Bulk raw read counts of human OVCAR5 samples were mapped to CellRanger's reference transcriptome refdata-cellranger-GRCh38-3.0.0 using version 0.9 of the Nextflow core kallisto pipeline (<https://github.com/cbcrg/kallisto-nf>) with kallisto 0.46.2. Read count distributions were a priori assessed using an awk script (as adapted from <https://www.biostars.org/p/243552/>). Transcript counts were collapsed to genes using the R package txImport (version 1.20.0) and Ensembl gene identifiers (IDs) were converted to HUGO gene names using a home-made lookup table generated from the CellRanger gtf file. Gene read counts were loaded into the R Seurat package (version 4.1.0), then TMM-library size normalized using edgeR (version 3.36.0) and finally additionally corrected using Seurat's regularized negative binomial model regression, in the same manner as the single cell data (SCT normalization⁴⁴). The last step had the effect of setting lowly expressed genes (fewer than 1 unit of TMM-normalized gene expression) to zero and also reduced the contrast in expression between highly and lowly expressed genes. Bulk sequencing data of murine NMM samples were aligned paired-end, strand and transcriptome aware, with hisat2⁴⁵ against GRCm38. Counts per gene were made using itreecount (<https://github.com/NKI-GCF/itreecount>) and annotated using ensembl gtf version 87.

Selection of cytokine-responsive genes

To identify cytokine-responsive genes for the OVCAR5 cell line using machine learning, we first devised a set of gene-characterizing features. A limma voom model⁴⁶ of the form $y \sim X_c + X_d$, where y reflects per gene expression levels, X s are design matrices, c reflect stimulus nature and concentration (one term for each combination of a tested stimulus and concentration), and d reflects stimulus exposure duration, was fitted to the 'OVCAR5 bulk RNA-seq reference data set' described in the section '*in vitro cytokine stimulation*' and Figure 5.1. Duration coefficients were included to absorb confounding duration gene expression dynamics that were independent of the nature of stimulus. Interaction terms between stimuli and durations were not included as the design matrices would not be full-rank, as we had exactly one replicate per experimental condition. We collected the following statistics:

1. Maximum t -statistics and effect sizes for the X_c terms were extracted from the fitted limma object.
2. A cytokine specificity score, capturing the relative response to either TNF- α or IFN- γ , which was computed as $|\beta_{\text{TNF-}\alpha}|/(|\beta_{\text{TNF-}\alpha}|+|\beta_{\text{IFN-}\gamma}|)$, where indicated coefficients reflect those of the highest tested concentrations of indicated cytokines (10 ng/ml for TNF- α and 100 ng/ml for IFN- γ). We have not noticed any genes responding more strongly to lower concentrations of a given stimulus, justifying this approach.
3. A set of genes responded more strongly to the combination of IFN- γ and TNF- α than expected based on their response to these stimuli in isolation. To describe this synergistic behavior, a com-

bined $\text{TNF-}\alpha$ plus $\text{IFN-}\gamma$ synergy gene set score was computed as

$$\beta_{\text{IFN-}\gamma+\text{TNF-}\alpha} / (\beta_{\text{TNF-}\alpha} + \beta_{\text{IFN-}\gamma}) - 1$$

where indicated coefficients again reflect the highest tested concentrations.

4. As some genes responded very strongly to stimuli but only at specific timepoints and such effects would get diluted in the aforementioned limma model (which can be interpreted as an estimated average effect of concentration across the different tested exposure durations), single time point statistics were additionally extracted. For each stimulus and time point, the log2 fold difference with the duration-matched unstimulated control sample was computed, resulting in 3 additional statistics ($\text{IFN-}\gamma$, $\text{TNF-}\alpha$ and $\text{IFN-}\gamma$ plus $\text{TNF-}\alpha$), as well as the maximum of these three statistics.

5. Genes for which one stimulus was consistently higher across the four tested timepoints appeared more informative than genes for which this was more variable. To describe this, the maximum number of exposure durations for which any given cytokine and concentration yielded the highest or lowest response was evaluated, yielding another two integral statistics ranging between 1 and 4.

6. With the same goal in mind, the Pearson correlation between all three pairs of consecutive timepoints (2 & 6, 6 & 12, 12 & 24) across the different stimuli was recorded and summarized by the median across the three different sets. High-scoring genes on this metric will have high similarity in the ordering of stimuli in terms of effectuated gene expression across exposure durations. Line plots (as in Figure S5.2c) of such genes will appear ordered, i.e., with a low degree of line crossing.

7. The maximum (log2-transformed) gene expression for each gene in the TMM-normalized expression data across samples was extracted, as well as the difference between the maximum and minimum gene expression values across samples.

The above features were computed for all 33,514 detected genes. As simple thresholding using these statistics gave suboptimal results (data not shown), the set of gene classifications was augmented from an original manually-classified set of genes ($n = 80$) in an iterative process of i) random forest model fitting on already classified genes, with gene class as the response variable and the aforementioned features as explanatory variables ii) class prediction for previously unclassified genes and iii) manual curation of these model predictions (Figure S5.2a). For step i), classification random forests were trained using the ranger engine (ranger package version 0.13.1) in R the tidymodels library (version 0.1.4) with importance set to 'impurity', using the aforementioned gene descriptive statistics. The mtry parameter was optimized using 3-fold cross validation on a random, unique sample of 75% of the already classified genes, leaving 25% of the genes for validation purposes. A final model was trained on all the training data using optimal hyperparameters. The model was trained to discern between the following gene classes: 'cytokine-unresponsive' (a gene for which none of the evaluated exposures leads to clearly elevated gene expression as judged by inspection of line-plot as in Figure S5.2c) and 'cytokine-responsive' (responsive to at least one cytokine stimulus). The latter class was

then subdivided into the following classes: ‘mono-responsive (a cytokine-responsive gene responding strongly to one of the two tested cytokines but not to the other, and for which the combination of TNF- α plus IFN- γ does not behave differently from the dominant cytokine), ‘synergistically-responsive’ (a cytokine-responsive gene responding *solely* to the combination of TNF- α plus IFN- γ and not to the individual cytokines), ‘other synergy’ (a cytokine-responsive gene that shows a moderate degree of stimulus synergy but for which individual cytokines also effectuate noticeable gene expression), ‘anti-synergy’ (a cytokine-responsive gene whose response to the combination of stimuli is weaker than to the sum of the individual stimuli), ‘lowly expressed’ (possibly too lowly expressed to be reliably detectable in single cell data) and ‘cytokine-responsive, other’ (responsive to cytokine exposure, but not fitting to any of the aforementioned class descriptions). Step ii) the final model was used to predict classes for all previously unclassified genes. Step iii) all ‘cytokine-responsive’ genes (and some ‘cytokine-unresponsive’ genes as well, to ensure the absence of false negative predictions) were inspected and predictions were adjusted where needed. After 10 iterations of model training and prediction curation, 612 cytokine-responsive genes were acquired (Figure S5.2d), at which point the yield of informative additional genes per additional cycle had slowed down to just a handful, suggesting nearly full extraction of all cytokine-responsive genes. For the purposes of this study, in which the ‘other synergy’, ‘anti-synergy’ and ‘lowly expressed’ classes are superfluous, genes belonging to these classes were next reclassified as ‘cytokine-responsive, other’ (Figure S5.2b). In addition, the ‘mono-responsive’ class was partitioned into ‘IFN- γ mono-responsive’ (genes with cytokine specificity score $\leq .5$) and ‘TNF- α mono-responsive’ (genes with cytokine specificity score $> .5$). Additionally, the following ‘time-informative’ gene sets were compiled by manual sub-selection from all 612 cytokine-responsive genes: IFN- γ late and TNF- α late (cytokine-responsive genes with a most pronounced response at 12-24h of stimulation), IFN- γ plateau (IFN- γ -responsive genes rising in expression until 6 hours, after which they remain constant) and TNF- α early (genes responding most pronounced at 2 hours of stimulation). TNF- α plateau, as well as IFN- γ early, mono-responsive genes could not be identified.

For the selection of genes for the IFN- γ and TNF- α gene sets for the murine NMM cell line, genes were prioritized using a simplification of the analysis done on the human OVCAR5 data. An identical limma model was fitted to the MNN bulk RNA-seq reference data set, a candidate gene list was generated based on the resulting moderated *t*-statistics and candidate genes were then manually filtered for cytokine-unresponsive genes by inspecting their ‘line’ plots (as in Figure 5.1a), resulting in a total of 134 mono-responsive genes. TGF- β -responsive genes in the Milo analysis (Figure 5.4e) were identified by filtering based on limma moderated *t*-statistics, with a limma model as described above applied (Figure S5.5a), resulting in one *t*-statistic for each of the evaluated single stimuli per gene. Genes were included if they were found mono-responsive to TGF- β , i.e., a) a *t*-statistic surpassing 4.2 for TGF- β and b) below 1 for all other evaluated single stimuli. Genes were then filtered for biological, as opposed to purely technical, expression variation using the modelGeneVar function in the R package scran (package version 1.20.1), which was called with log2(cpm + 1) transformed data and with exposure duration meta information as the function’s argument to ‘block’. Genes were required

to have $\text{bio} > 0$ and $\text{FDR} \leq 10^{-7}$, resulting in a total of 73 TGF- β mono-responsive genes.

Human OVCAR5 single cell sequencing data preprocessing

Cell Ranger UMI count data was loaded into R Seurat objects (version 4.1.0), and cells with less than 1,000 detected RNA features were filtered out. Across single cell sequencing experiments, either a single or two hashtag bar codes per sample (i.e., experimental condition) were employed, the latter to allow inclusion of a larger number of samples than the number of available hashtagging antibodies. Sample assignment for experiments employing single hashtagging was done using Seurat's (package version 4.1.0) HTODemux functionality on the CLR-normalized barcode hashtag data using default settings. Sample assignment for experiments employing double hashtagging was done using a custom functionality. First, HTOs were CLR normalized using Seurat's `NormalizeData`. Next, the product of normalized hashtag counts was computed for each theoretically possible combination of two different hashtag antibodies. Cells were assigned to the sample corresponding to the highest product of CLR-normalized hashtag counts. Low-confidence assignments were then filtered out based on the fold difference between the dominant and second to dominant hashtag combination. A threshold value of 2 for this statistic was picked by comparing cells for which the dominant combination of hashtags was expected (i.e., a combination included in the experimental design) and those for which it was not. Cells were additionally filtered for a maximum mitochondrial content of 30%. SCT total UMI count normalization was performed on the remaining set of cells⁴⁴. We next identified the most variable features in the experiment (Seurat's `FindVariableFeatures` with default settings) and performed principal component analysis with 10 principal components (sufficient, as indicated by scree plots, obtained using Seurat's `RunPCA`, default settings) and a UMAP (Seurat's `RunUMAP`, default settings) over the principal component scores. Next, outlying clusters in the UMAP were automatically identified using density-based clustering on the UMAP cell coordinates with DBSCAN (fpc package, version 2.2.9) using parameters: $\text{eps} = .6$ and $\text{MinPts} = 6$. Small DBSCAN clusters (less than $1/(5c)$ cells, where c is the number of detected DBSCAN clusters) were then marked as outlying clusters and removed from downstream analyses. Using differential gene expression (DGE) analysis (Seurat's `FindMarkers`) and GSEA with (R 'fgsea' package, version 1.18.0) with REACTOME pathways⁴⁷ on the logFC-ranked list of differentially expressed genes (filtered first FDR-adjusted p-value $\leq .1$) between the outlying cluster and the main body of cells (i.e., the composite of non-outlying clusters), we identified these outlying clusters (0% -2.7% of cells across experiments) to likely consist of keratinocytes and/or fibroblasts, characterized by high *KRT81*, *S100A9* and *LDHB* counts. SCT normalization was redone on the remaining cells. Using the intersection of cytokine-responsive genes and detectable genes for each experiment, PCA was recomputed with 10 PCs and UMAPs were recomputed based on the PC scores.

Murine NMM single cell sequencing data preprocessing

The murine NMM data were preprocessed identically to the human OVCAR5 single cell data. Inspecting the initial UMAP, we noticed two clusters of cells. Characterizing the smaller cluster (DBSCAN cluster 2, 21.5% of cells), we noticed a 7.2-fold difference in mean UMIs between clusters (47,753 versus 6,654 mean UMI), suggestive of the presence of dying cells and/or cell fragments in the second cluster. Cells in this cluster were then removed, and SCT normalization, PCA and UMAP computation was repeated on the remaining cells.

Heatmap visualization of bulk RNA-seq data

TMM and SCT normalized data were log2-transformed and subsequently Z-scaled across genes/features. Unless indicated otherwise, rows (genes) and columns (samples) were subsequently clustered using complete linkage hierarchical clustering (hclust function in R), using Spearman correlation distance (computed as $1 - c$, where c is the correlation between two samples) for genes (rows) and Euclidean distance for samples (columns).

Gene set scores for single cell data

For both OVCAR5 and NMM cell lines, gene set scores were computed by summing SCT-normalized expression values of a gene set's member genes. Where indicated, gene set scores were normalized to the distribution of gene set scores of a relevant, duration-matched control condition with the following transformation: $f(x) = x - m(x_c)/IQR(x_c)$, where m represents the median, x_c is a vector of scores for the same gene set for a reference condition and $IQR(x_c)$ is the interquartile range in that reference population.

Quantification of separability of experimental conditions using gene set scores

AUROC values in main text and figures represent the area under the receiver operator curve from a binary classifier aiming to separate the two indicated experimental conditions with the indicated gene set score as the sole explanatory variable. These scores were computed using the `roc_auc_vec` function in the R `yardstick` package (version 0.0.9). AUROC values range between 0.5 (signifying no separation between experimental conditions) and 1 (signifying complete separation between experimental conditions). To distinguish more than two classes at a time, as in Figure 5.2b, we trained SVM models using the Gaussian kernel on 75% of the data, leaving the remaining 25% as validation data, with the R `tidymodels` framework (package version 0.1.4). Hyperparameters `cost` and `rbf_sigma` were optimized in a 10-fold cross validation within the training data. Optimal parameters were then used to train a final model on the full set of training data, of which the performance was sub-

sequently evaluated on the validation data. Included confusion matrices show the correspondence between actual and predicted class labels in the validation data, using a model trained on independent data, ignoring the fact that single cell transcriptomes in the training and validation sets were jointly preprocessed.

Milo neighborhood analysis of single cell data

A PCA of the murine single cell data was computed using the 2,000 most variable genes (as identified using Seurat's FindVariableGenes with default settings) and 20 principal components. A KNN graph of the data was constructed on the principal component scores with Milo's (package version 1.0.0) buildGraph function using $k = 10$ and $d = 20$. Neighborhoods were defined using makeNhoods, with refined= TRUE, prop = 0.1 and identical settings for k and d as aforementioned. Neighborhoods were tested for differential abundance of the three experimental conditions using the testNhoods function, with a design matrix that was obtained from a metadata table with the following R model formula: $\sim 1 + \text{experimental condition}$, resulting in an intercept term and a regression coefficient for the second and third experimental conditions. In visualizations of the KNN graph, effect sizes (logFC DA) associated with a Spatial FDR > 0.05 were whitened.

Deconvolution analysis of single cell neighborhood expression

A gene expression matrix of g genes (rows) by n single cell neighborhoods (SCNs, columns) in linear space was extracted using the Milo nhooExpression class method, TMM-transformed to normalize library sizes and finally log10-transformed, i.e., $M_N = \log_{10}(TMM(X) + 1)$, wherein X is the output of nhooExpression and M_N is the processed matrix of SCN expression. Similarly, a $g \times 28$ reference profile matrix M_R consisting of bulk reference RNASeq libraries with the 28 conditions/columns unstimulated 2h -unstimulated 24h (i.e., no stimulation for 2 hours followed by no stimulation for another 24 hours), unstimulated 2h -10 ng/ml TGFb 24h, unstimulated 2h -100 ng/ml IFNy 24h, 10 ng/ml TGFb 2h -unstimulated 24h, 10 ng/ml TGFb 2h -100 ng/ml IFNy 24h, 100 ng/ml IFNy 2h -unstimulated 24h, 100 ng/ml IFNy 2h -10 ng/ml TGFb 24h, unstimulated 6h -unstimulated 24h, unstimulated 6h -10 ng/ml TGFb 24h, unstimulated 6h -100 ng/ml IFNy 24h, 10 ng/ml TGFb 6h -unstimulated 24h, 10 ng/ml TGFb 6h -100 ng/ml IFNy 24h, 100 ng/ml IFNy 6h -unstimulated 24h, 100 ng/ml IFNy 6h -10 ng/ml TGFb 24h, unstimulated 12h -unstimulated 24h, unstimulated 12h -10 ng/ml TGFb 24h, unstimulated 12h -100 ng/ml IFNy 24h, 10 ng/ml TGFb 12h -unstimulated 24h, 10 ng/ml TGFb 12h -100 ng/ml IFNy 24h, 100 ng/ml IFNy 12h -unstimulated 24h, 100 ng/ml IFNy 12h -10 ng/ml TGFb 24h, unstimulated 24h -unstimulated 24h, unstimulated 24h -10 ng/ml TGFb 24h, unstimulated 24h -100 ng/ml IFNy 24h, unstimulated 24h -100 ng/ml IFNy 10 ng/ml TGFb 24h, 10 ng/ml TGFb 24h -unstimulated 24h, 10 ng/ml TGFb 24h -100 ng/ml IFNy 24h, 100 ng/ml IFNy 24h -unstimulated 24h, 100 ng/ml IFNy 24h -10 ng/ml TGFb 24h in the columns and genes in the rows) were TMM-normalized and then similarly log10-transformed. The

matrices were identically row-ordered (genes), with as rows the intersection of detected genes in the two unprocessed source matrices ($g = 14,620$) such that they predominantly consisted of genes that are minimally or not responsive to the cytokines of interest.

To estimate a SCN's stimulus exposure, its transcriptome was modeled as a linear combination of the columns (samples) of M_R using lasso l_1 -penalized multivariate regression with the regression coefficients (β) constrained to be larger than or equal to 0, such that they can be interpreted as mixing weights. The `cv.glmnet` function from the R package `glmnet` was used for this regression⁴⁸, as well as to optimize the lambda penalty for regression coefficients with arguments: `lower.limits=c(0)`, `family='gaussian'`, `alpha=.99`. The mean squared error between the original transcriptome (y) and reconstructed transcriptome ($x\beta$), i.e., $E[(y - x\beta)^2]$, was then extracted from the `cv.glmnet` object with lambda set to the value that minimized the cross validation error. To assess the importance of a subset of reference samples (columns) c for the reconstruction of any particular SCN, the reconstruction error was recomputed with the remaining columns of M_R after having removed samples/columns c . In this, a large increase in error can be interpreted as an indication that the samples c contain co-variation in gene expression that cannot be accommodated by any of the remaining reference samples. To quantify the possibility that reconstruction error was simply raised by providing the optimization algorithm a smaller set of basis vectors to work with (i.e., by limiting the span of M_R), permutation testing was employed. Specifically, the reconstruction error was first computed with 1,000 random selections of s reference profiles, where s is the number of reference profiles that were *not* associated with the reference profiles (columns) of which the importance is assessed. For example, to assess the importance of TGF- β stimulation, s is the number of profiles obtained from experiments where no TGF- β stimulation was applied (i.e. only unstimulated and IFN- γ stimulated samples, $s = 11$, Figure 5.4g, left). Similarly, to assess the importance of IFN- γ stimulation, s represents the number of profiles obtained from experiments where no IFN- γ stimulation was applied (i.e., only unstimulated and TGF- β -stimulated samples, $s = 11$, Figure 5.4g, right). The quantile of the observed error in the distribution of permutation errors was then acquired using the R function `ecdf`. SCN reconstruction was finally labeled as worse (or better) than expected if the observed error was in the 97.5st or higher (or 2.5th or lower) percentile of the permutation distribution.

Software and code availability

All preprocessing and analyses were done in R (version 4.1.0), frequently employing tidyverse packages (version 1.3.1). Row-or column-annotated heatmaps were made using the `ComplexHeatmap` package (version 2.9.3). Remaining plots were made using the `ggplot2` package (version 3.3.5). All data processing and time-intensive computation was done inside of the R targets pipelining system⁴⁹. Code and notebooks can be accessed at www.github.com/slagtermaarten/cyto_inference.

Acknowledgements

We thank K. Bresser, A. M. van der Leun, L. Kok and S. Mourragui for input and valuable discussions. We thank staff of the NKI Genomics Core facility for technical support, along with the NKI Research High Performance Computing, Animal Intervention, and Flow Cytometry facilities. This work was supported by a Boehringer Ingelheim Fonds PhD Fellowship (to M. E. H.), ERC AdG SENSIT, grant agreement ID 742259 (to T. N. M. S.) and institutional funding of the Netherlands Cancer Institute by the Dutch Cancer Society.

Author contributions

M.E.H. conceived the study together with M.S. and T.N.S, performed wet lab experiments and wrote the manuscript with M.S., L.F.A.W. and T.N.S.. M.S. performed and interpreted bioinformatic analyses. J.U. & M.T. aided in wet lab experiments. N.S. performed exploratory data analysis. R.K and M.N, designed and performed sequencing experiments and I.R. & R.J.C.K performed data preprocessing. L.F.A.W. and T.N.S. supervised bioinformatic analyses, and T.N.S. supervised wet-lab experiments. All authors have read and approved the manuscript.

Declaration of interests

L.F.A.W. received project funding for unrelated work from Bristol-Myers-Squibb. T.N.S. is advisor for Allogene Therapeutics, Asher Bio, Merus, Neogene Therapeutics, and Scenic Biotech; is a stockholder in Allogene Therapeutics, Asher Bio, Cell Control, Celsius, Merus, and Scenic Biotech; and is venture partner at Third Rock Ventures, all outside of the current work.

Bibliography

1. Giladi, A. *et al.* Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat Biotechnol* **38**, 629-637, doi:10.1038/s41587-020-0442-2 (2020).
2. Altan-Bonnet, G. & Mukherjee, R. Cytokine-mediated communication: a quantitative appraisal of immune complexity. *Nat Rev Immunol* **19**, 205-217, doi:10.1038/s41577-019-0131-x (2019).
3. Raskov, H., Orhan, A., Christensen, J. P. & Gogenur, I. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer* **124**, 359-367, doi:10.1038/s41416-020-01048-4 (2021).
4. Rosenberg, S. A. *et al.* Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin Cancer Res* **17**, 4550-4557, doi:10.1158/1078-0432.CCR-11-0116 (2011).

-
5. Dudley, M. E. *et al.* Randomized selection design trial evaluating CD8+-enriched versus unselected tumor-infiltrating lymphocytes for adoptive cell therapy for patients with melanoma. *J Clin Oncol* **31**, 2152-2159, doi:10.1200/JCO.2012.46.6441 (2013).
 6. Morotti, M. *et al.* Promises and challenges of adoptive T-cell therapies for solid tumours. *Br J Cancer* **124**, 1759-1776, doi:10.1038/s41416-021-01353-6 (2021).
 7. Mojic, M., Takeda, K. & Hayakawa, Y. The Dark Side of IFN-gamma: Its Role in Promoting Cancer Immune evasion. *Int J Mol Sci* **19**, doi:10.3390/ijms19010089 (2017).
 8. Castro, F., Cardoso, A. P., Goncalves, R. M., Serre, K. & Oliveira, M. J. Interferon-Gamma at the Crossroads of Tumor Immune Surveillance or Evasion. *Front Immunol* **9**, 847, doi:10.3389/fimmu.2018.00847 (2018).
 9. Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol* **8**, 958-969, doi:10.1038/nri2448 (2008).
 10. Jorgovanovic, D., Song, M., Wang, L. & Zhang, Y. Roles of IFN-gamma in tumor progression and regression: a review. *Biomark Res* **8**, 49, doi:10.1186/s40364-020-00228-x (2020).
 11. Parameswaran, N. & Patial, S. Tumor necrosis factor-alpha signaling in macrophages. *Crit Rev Eukaryot Gene Expr* **20**, 87-103, doi:10.1615/critreveukargeneexpr.v20.i2.10 (2010).
 12. Trevejo, J. M. *et al.* TNF-alpha -dependent maturation of local dendritic cells is critical for activating the adaptive immune response to virus infection. *Proc Natl Acad Sci U S A* **98**, 12162-12167, doi:10.1073/pnas.211423598 (2001).
 13. Braumuller, H. *et al.* T-helper-1-cell cytokines drive cancer into senescence. *Nature* **494**, 361-365, doi:10.1038/nature11824 (2013).
 14. Montfort, A. *et al.* The TNF Paradox in Cancer Progression and Immunotherapy. *Front Immunol* **10**, 1818, doi:10.3389/fimmu.2019.01818 (2019).
 15. Wang, W. *et al.* CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* **569**, 270-274, doi:10.1038/s41586-019-1170-y (2019).
 16. Kammertoens, T. *et al.* Tumour ischaemia by interferon-gamma resembles physiological blood vessel regression. *Nature* **545**, 98-102, doi:10.1038/nature22311 (2017).
 17. Briesemeister, D. *et al.* Tumor rejection by local interferon gamma induction in established tumors is associated with blood vessel destruction and necrosis. *Int J Cancer* **128**, 371-378, doi:10.1002/ijc.25350 (2011).
 18. Zhang, B., Karrison, T., Rowley, D. A. & Schreiber, H. IFN-gamma- and TNF-dependent bystander eradication of antigen-loss variants in established mouse cancers. *J Clin Invest* **118**, 1398-1404, doi:10.1172/JCI33522 (2008).
 19. Spiotto, M. T., Rowley, D. A. & Schreiber, H. Bystander elimination of antigen loss variants in established tumors. *Nat Med* **10**, 294-298, doi:10.1038/nm999 (2004).

20. Spiotto, M. T. & Schreiber, H. Rapid destruction of the tumor microenvironment by CTLs recognizing cancer-specific antigens cross-presented by stromal cells. *Cancer Immun* **5**, 8 (2005).
21. Huse, M., Lillemeier, B. F., Kuhns, M. S., Chen, D. S. & Davis, M. M. T cells use two directionally distinct pathways for cytokine secretion. *Nat Immunol* **7**, 247-255, doi:10.1038/ni1304 (2006).
22. Kupfer, A., Mosmann, T. R. & Kupfer, H. Polarized expression of cytokines in cell conjugates of helper T cells and splenic B cells. *Proc Natl Acad Sci U S A* **88**, 775-779, doi:10.1073/pnas.88.3.775 (1991).
23. Sanderson, N. S. *et al.* Cytotoxic immunological synapses do not restrict the action of interferon-gamma to antigenic target cells. *Proc Natl Acad Sci U S A* **109**, 7835-7840, doi:10.1073/pnas.1116058109 (2012).
24. Hoekstra, M. E., Vijver, S. V. & Schumacher, T. N. Modulation of the tumor micro-environment by CD8(+) T cell-derived cytokines. *Curr Opin Immunol* **69**, 65-71, doi:10.1016/j.coi.2021.03.016 (2021).
25. Perona-Wright, G., Mohrs, K. & Mohrs, M. Sustained signaling by canonical helper T cell cytokines throughout the reactive lymph node. *Nat Immunol* **11**, 520-526, doi:10.1038/ni.1866 (2010).
26. Ariotti, S. *et al.* T cell memory. Skin-resident memory CD8(+) T cells trigger a state of tissue-wide pathogen alert. *Science* **346**, 101-105, doi:10.1126/science.1254803 (2014).
27. Schenkel, J. M. *et al.* T cell memory. Resident memory CD8 T cells trigger protective innate and adaptive immune responses. *Science* **346**, 98-101, doi:10.1126/science.1254536 (2014).
28. Muller, A. J. *et al.* CD4+ T cells rely on a cytokine gradient to control intracellular pathogens beyond sites of antigen presentation. *Immunity* **37**, 147-157, doi:10.1016/j.immuni.2012.05.015 (2012).
29. Thibaut, R. *et al.* Bystander IFN-gamma activity promotes widespread and sustained cytokine signaling altering the tumor microenvironment. *Nat Cancer* **1**, 302-314, doi:10.1038/s43018-020-0038-2 (2020).
30. Hoekstra, M. E. *et al.* Long-distance modulation of bystander tumor cells by CD8(+) T cell-secreted IFNgamma. *Nat Cancer* **1**, 291-301, doi:10.1038/s43018-020-0036-4 (2020).
31. Beck, R. J., Slagter, M. & Beltman, J. B. Contact-Dependent Killing by Cytotoxic T Lymphocytes Is Insufficient for EL4 Tumor Regression In Vivo. *Cancer Res* **79**, 3406-3416, doi:10.1158/0008-5472.CAN-18-3147 (2019).
32. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
33. Buttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat Methods* **16**, 43-49, doi:10.1038/s41592-018-0254-1 (2019).

-
34. Fukunaga, R., Sokawa, Y. & Nagata, S. Constitutive production of human interferons by mouse cells with bovine papillomavirus as a vector. *Proc Natl Acad Sci U S A* **81**, 5086-5090, doi:10.1073/pnas.81.16.5086 (1984).
 35. Savan, R., Ravichandran, S., Collins, J. R., Sakai, M. & Young, H. A. Structural conservation of interferon gamma among vertebrates. *Cytokine Growth Factor Rev* **20**, 115-124, doi:10.1016/j.cytogfr.2009.02.006 (2009).
 36. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol* **40**, 245-253, doi:10.1038/s41587-021-01033-z (2022).
 37. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).
 38. Garnier, L. *et al.* IFN-gamma-dependent tumor-antigen cross-presentation by lymphatic endothelial cells promotes their killing by T cells and inhibits metastasis. *Sci Adv* **8**, eabl5162, doi:10.1126/sciadv.abl5162 (2022).
 39. Boutilier, A. J. & Elswa, S. F. Macrophage Polarization States in the Tumor Microenvironment. *Int J Mol Sci* **22**, doi:10.3390/ijms22136995 (2021).
 40. Wen, F. Q. *et al.* Interleukin-4- and interleukin-13-enhanced transforming growth factor-beta2 production in cultured human bronchial epithelial cells is attenuated by interferon-gamma. *Am J Respir Cell Mol Biol* **26**, 484-490, doi:10.1165/ajrcmb.26.4.4784 (2002).
 41. Burd, C. E. *et al.* Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov* **4**, 1418-1429, doi:10.1158/2159-8290.CD-14-0729 (2014).
 42. Vredevoogd, D. W. *et al.* Augmenting Immunotherapy Impact by Lowering Tumor TNF Cytotoxicity Threshold. *Cell* **178**, 585-599 e515, doi:10.1016/j.cell.2019.06.014 (2019).
 43. van Rooij, N. *et al.* Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol* **31**, e439-442, doi:10.1200/JCO.2012.47.7521 (2013).
 44. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).
 45. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915, doi:10.1038/s41587-019-0201-4 (2019).
 46. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

47. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687-D692, doi:10.1093/nar/gkab1028 (2022).
48. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
49. Landau, W. M. The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *. Journal of Open Source Software* **6**(57), doi: 10.21105/joss.02959 (2021).

Supplemental Figures

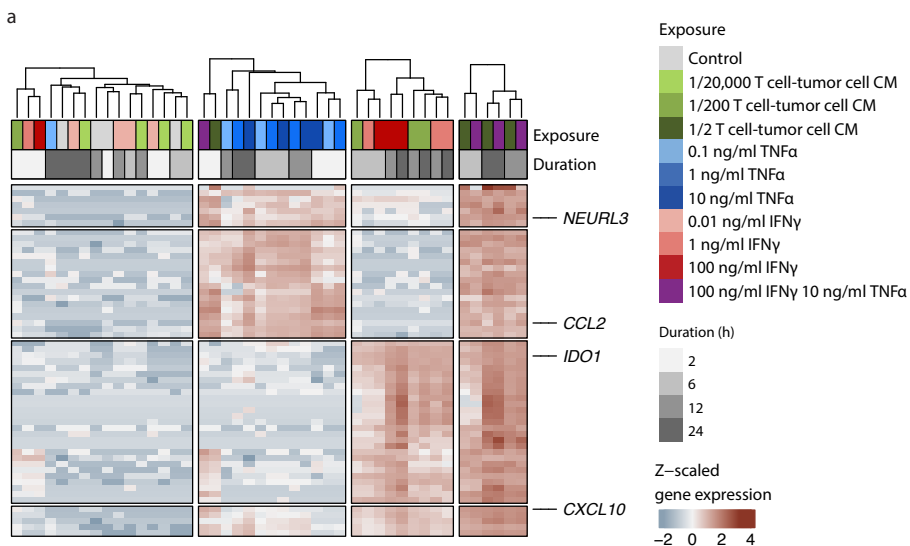


Figure S5.1: Gene expression profiles of tumor cells exposed to IFN- γ plus TNF- α or T cell - tumor cell co-culture medium.

OVCAR5 tumor cells were stimulated with recombinant IFN- γ , TNF- α , or IFN- γ plus TNF- α (as in Figure 5.1b), or with tissue culture medium (CM) obtained from co-cultures of CDK4_{R>L} Ag⁺ OVCAR5 tumor cells and CDK4_{R>L}-specific CD8⁺ T cells, for the indicated times and concentrations. Unsupervised hierarchical clustering of bulk RNAseq data, with genes filtered for a variance of at least 0.4 across the included samples in the library size normalized but unscaled representation of the data. Note that T cell-tumor co-culture medium-exposed samples co-cluster with TNF- α plus IFN- γ -stimulated samples.

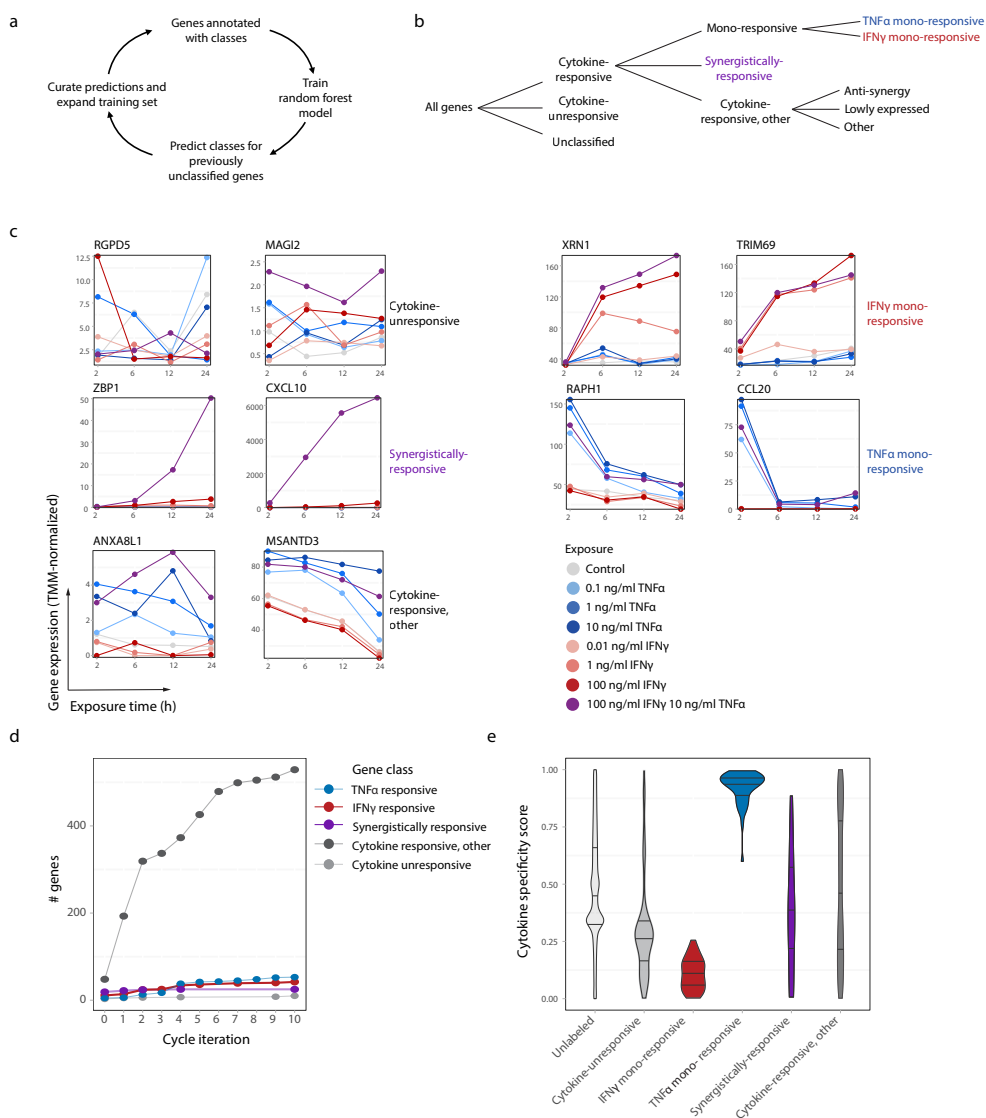


Figure S5.2: Model prediction and curation cycle to identify cytokine-responsive genes.

Starting with a small set of handpicked genes, a machine learning model trained on this initial seed set was used to efficiently identify additional cytokine-responsive genes of the indicated classes. Line plots (as shown in c) of each gene predicted to be cytokine-responsive were manually inspected and gene classes were corrected where necessary. **a** Hierarchy of gene class assignment. Members of the ‘TNF- α mono-responsive’ (blue) and ‘IFN- γ mono-responsive’ (red) gene classes were used to calculate cytokine gene set scores. Synergistically-responsive genes (purple), responding solely to the combination of TNF- α plus IFN- γ and not to the individual cytokines, form the ‘synergy gene set’ used to calculate synergy gene set scores.

b RNA expression profiles of two randomly selected members of the indicated gene classes, as described in b. Expression profiles depict bulk RNAseq data from OVCAR5 cells exposed *in vitro* to the indicated concentration of recombinant IFN- γ , TNF- α or IFN- γ plus TNF- α , for the indicated duration. **c** Gene counts per class after each prediction and curation cycle. Flattening of the curves with increasing numbers of cycles suggests that most cytokine-responsive genes have been identified. **d** The cytokine specificity score (See Methods section

'Selection of cytokine-responsive genes'), quantifying differential transcriptional responsiveness to either TNF- α (value of 1) or IFN- γ (value of 0), for members of the indicated gene classes.

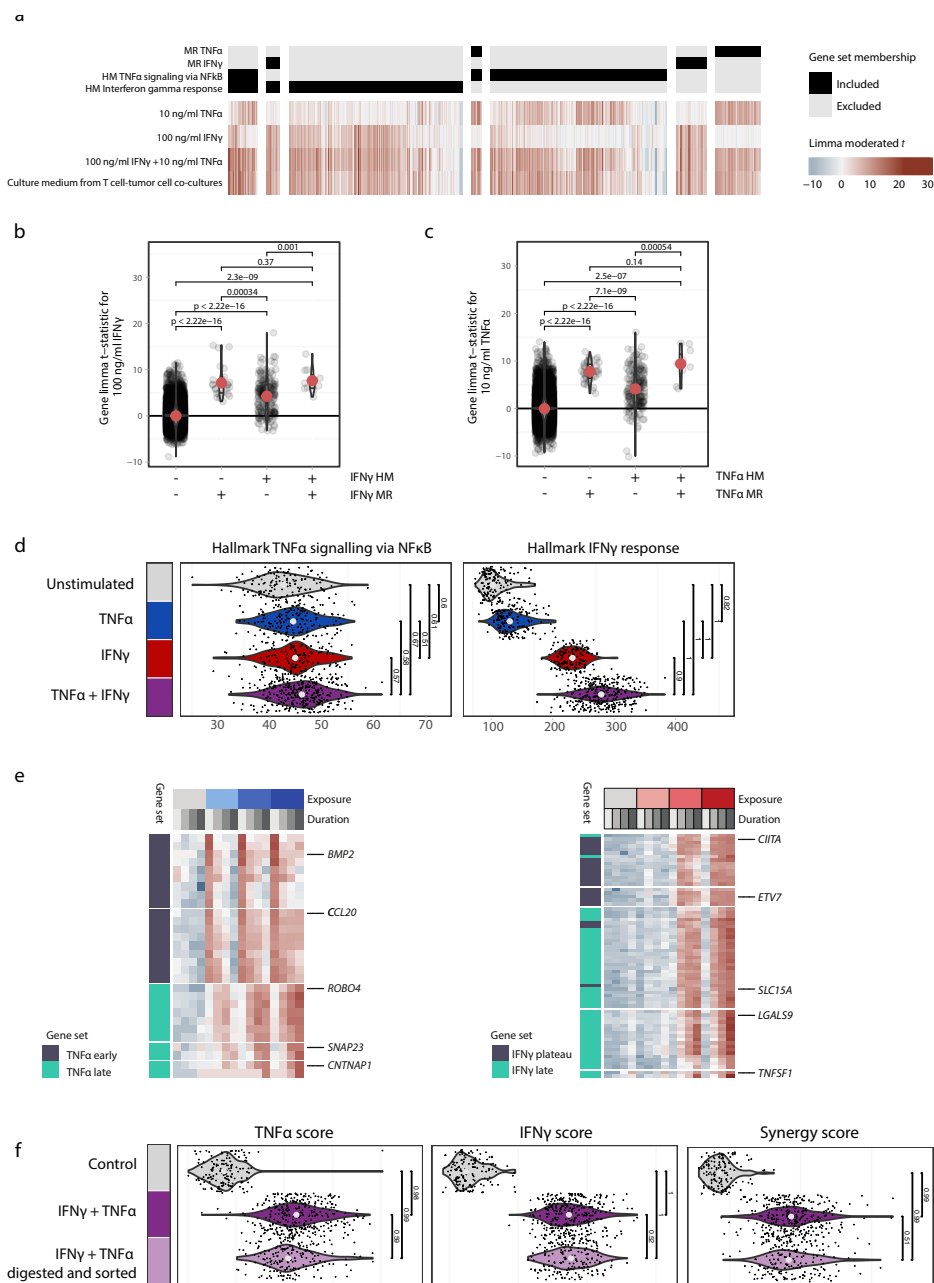


Figure S5.3: Validation and characterization of cytokine-responsive genes in bulk RNA-seq data.
a Limma moderated t -statistics for activating concentrations of cytokines (100ng/ml IFN- γ , 10ng/ml TNF- α) in the OVCAR5 bulk RNAseq reference data set (See methods section 'In vitro cytokine stimulation' and Figure

5.1). Depicted genes are present in one of the mono-responsive (MR) gene sets identified in this work and/or in one or both of the Hallmark 'Interferon gamma response' and 'TNF- α signaling via NFkB' gene sets³². Note that whereas MR genes were selected based on their response to exactly one stimulus in OVCAR5 cells (among the cytokines tested), the majority of Hallmark genes did not qualify as MR, frequently responding to multiple cytokines, or not responding to the purported stimulus in OVCAR5 cells.

b Comparison of limma moderated *t*-statistics between genes that are in- or excluded from the Hallmark Interferon gamma response (HM) gene set and in- or excluded from the OVCAR5-specific IFN- γ MR set. Black dots represent individual genes; red dots denote medians. Displayed comparisons are Wilcoxon rank sum test *p*-values.

c As in **b**, but with the Hallmark 'TNF- α signaling via NFkB' and OVCAR5 TNF- α MR gene sets.

d As in Figure 5.1e, but with the Hallmark gene sets instead of the OVCAR5 customized ones. The 'TNF- α signaling via NFkB' set shows virtually no response to TNF- α stimulation, whereas the 'Interferon gamma response' responds more strongly to the combination of TNF- α and IFN- γ than it does to IFN- γ alone.

e Heatmaps of time-informative genes for TNF- α (left panel) and IFN- γ (right panel). 'Early' genes reach maximal expression after 2 hours of cytokine exposure, with reduced expression at later timepoints. 'Plateau' genes reach maximal expression at 6 hours of cytokine exposure, with relatively constant expression at later timepoints. 'Late' genes show a continuing increase in expression up to 24 hours of cytokine exposure.

f Violin plots depicting gene set scores of *in vitro* cultured and cytokine stimulated (100 ng/ml IFN- γ and 10 ng/ml TNF- α for 24h) OVCAR5 cells that were or were not exposed to single cell digestion and fluorescence activated cell sorting. Dots represent gene set scores of individual cells, violins represent densities of score distributions. Values represent AUROC values, denoting the separability of cells under the indicated conditions (with 0.5 indicating no separation and 1 indicating a complete separation).

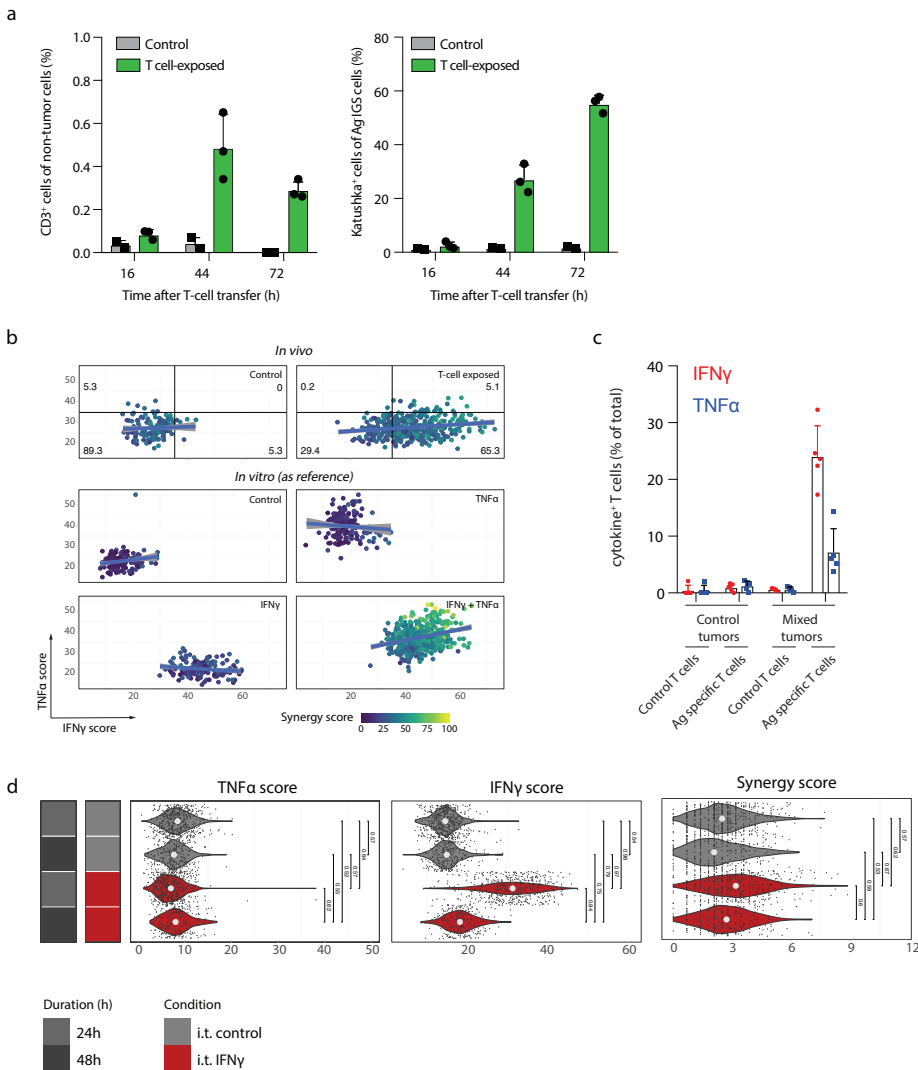


Figure S5.4: Quantification of T-cell infiltration and functionality *in vivo*.

a NSG- $\beta 2m^{-/-}$ mice were injected subcutaneously with a mixture of 10% CDK4_{R>L} antigen expressing (Ag⁺) OVCAR5 cells and 90% Ag⁻ CFP⁺ bystander OVCAR5 tumor cells carrying a fluorescence based IFN- γ sensing (IGS) reporter. After tumor establishment, tumors were treated by intravenous injection of PBS (control) or CDK4_{R>L} specific CD8⁺ T cells, and tumors were harvested at the indicated timepoint after treatment. Left panel: percentage of CD3⁺ cells of non-tumor cells, as measured by flow cytometry. Right panel: fraction of Ag⁻ bystander tumor cells with an activated IGS reporter (katushka⁺) from the total bystander tumor cell population, as measured by flow cytometry.

b Upper two panels: scatter plots of TNF- α and IFN- γ gene set scores from OVCAR5 cells derived from control and T cell-exposed tumors. The horizontal and vertical lines demarcate the 95th percentiles of the *in vivo*

control condition for the TNF- α gene set score and IFN- γ gene set score, respectively. Bottom four panels: scatter plots of TNF- α and IFN- γ gene set scores from OVCAR5 cells that have been stimulated *in vitro* with recombinant cytokines, as described in Figure 5.1, serving as a reference for the two top panels. Note that the synergy gene set score (color scale) is selectively elevated in cells that show high expression of both the TNF- α and IFN- γ mono-responsive gene sets.

c Tumor-infiltrated T cells retain the capacity to produce IFN- γ and TNF- α . NSG- $\beta 2m^{-/-}$ mice were injected subcutaneously with either a mixture of 10% antigen expressing (Ag⁺) OVCAR5 cells and 90% Ag⁻ CFP⁺ bystander OVCAR5 tumor cells (mixed tumors) or with Ag⁻ CFP⁺ tumor cells (control tumors). After tumor establishment, mice were treated by intravenous injection of a mixture of control CD8⁺ T cells (40%) and specific CD8⁺ T cells (60%), and tumors were harvested 44h after treatment. Digested tumors were subsequently cultured *in vitro* in the presence of Ag⁻ CFP⁺ tumor cells (for digests from control tumors) or Ag⁺ tumor cells (for digests from Ag⁺/Ag⁻ mixed tumors) for 3 hours in the presence of Golgi-plug to evaluate capacity for continued cytokine production. Subsequently, cells were stained for intracellular IFN- γ and TNF- α and analyzed by flow cytometry. Each dot represents T cells derived from one tumor, bar graphs show mean of the indicated groups + SD, $n=5$ mice per group. Data of one experiment is depicted. Note that specific CD8⁺ T cells derived from mixed tumors retain the capacity to produce TNF- α and IFN- γ .

d Gene set scores of OVCAR5 tumor cells obtained from subcutaneous OVCAR5 tumors isolated 24h or 48h after intratumoral injection of IFN- γ or PBS (control). Dots represent gene set scores of individual cells, violins represent densities of score distributions. Values represent AUROC values, denoting the separability of cells under the indicated conditions (with 0.5 indicating no separation and 1 indicating a complete separation). Note that a single injection of IFN- γ leads to a transient increase in IFN- γ gene set scores that is nearly reversed at 48h.

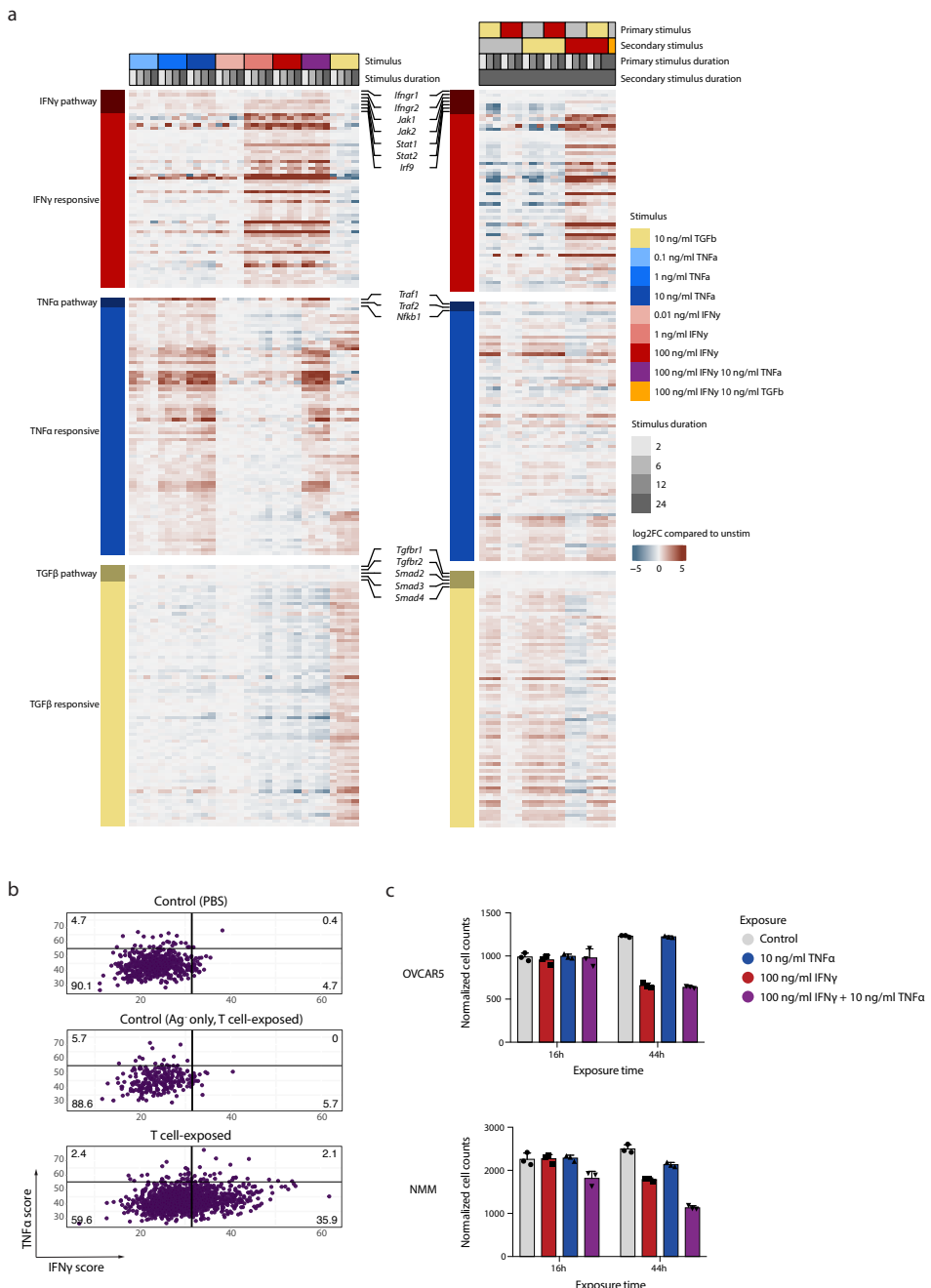


Figure S5.5: Characterization of cytokine-responsiveness in a syngeneic tumor model.

a Heatmap of bulk RNAseq data of murine NMM melanoma cells exposed *in vitro* to the indicated cytokines. Left and right show two separate RNAseq experiments, with different (combinations of) stimuli. Depicted 'Responsive' genes were manually selected for mono-responsiveness to either IFN- γ , TNF- α or TGF- β . Shown are the ratios between expression in cytokine stimulated and control condition for the same duration of time. Control conditions are hence not shown. **b** Scatter plot of IFN- γ vs. TNF- α gene set scores of data described in

Figure 5.4a-b, in which every point represents a cell. Plots are partitioned based on the 95th percentiles of the first control condition (top panel). **c** Quantification of cytokine mediated growth inhibition of OVCAR5 and NMM bystander tumor cells *in vitro*. Relative cell counts of CFP⁺ OVCAR5 (left) and CFP⁺ NMM (right) tumor cells after incubation in the absence or presence of recombinant IFN- γ TNF- α or IFN- γ *plus* TNF- α for the indicated time periods. Cells were analyzed by flow cytometry and cell counts were normalized to counting beads. Bar graph shows mean of 3 technical replicates. Representative data of two independent experiments are depicted. The cell inhibition that is observed upon IFN- γ exposure (in either the presence or absence of TNF- α) for OVCAR5 at 44 hours indicates that analysis of single cell RNA-seq data may somewhat (up to a factor of 2) underestimate IFN- γ sensing at this timepoint for this cell line (as 'sensing' cells are at a disadvantage). The cell inhibition that is observed upon IFN- γ exposure and that is enhanced by additional TNF- α exposure for NMM at 44 hours indicates that analysis of single cell RNA-seq data may somewhat underestimate IFN- γ (up to a factor of 1.3) and IFN- γ -TNF- α (up to a factor of 2) sensing at this timepoint for this cell line. Note that such cell inhibition does not influence estimates at the 16hr timepoint.

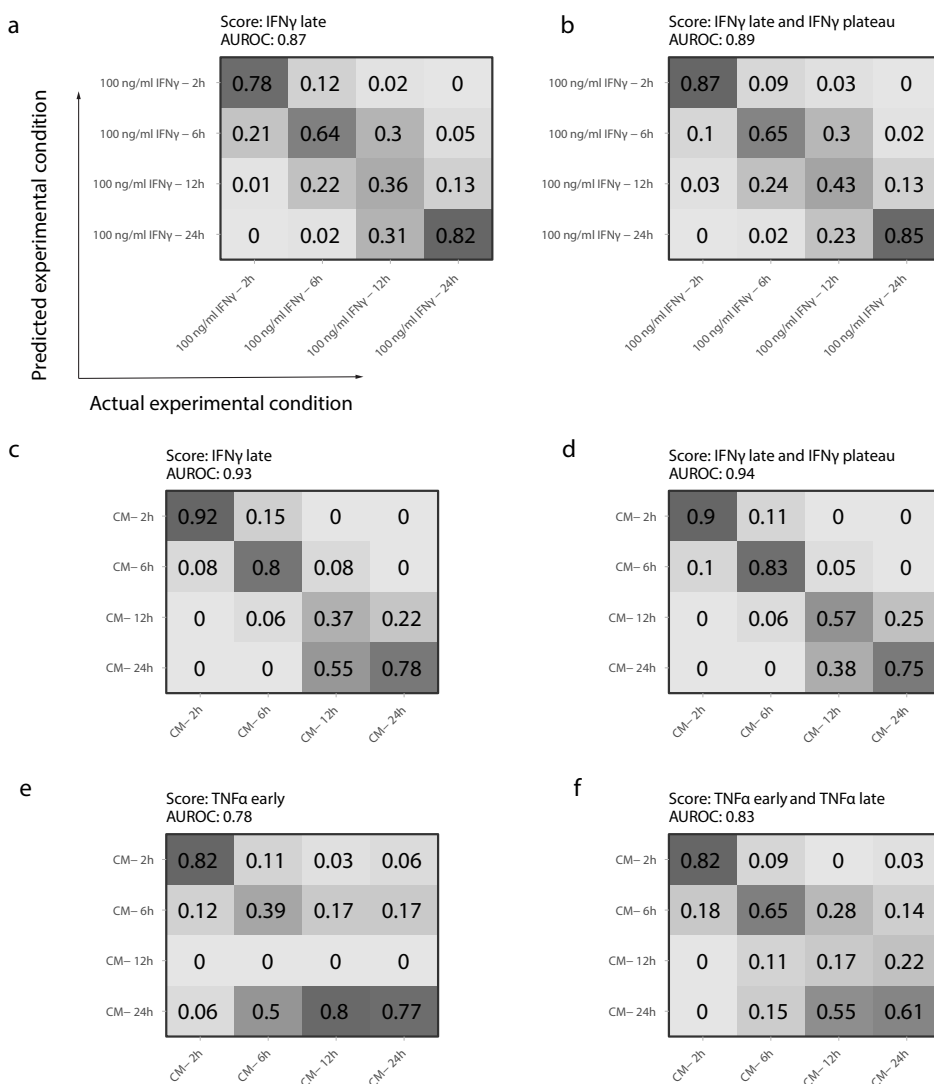


Figure S5.6: Quantification of exposure duration separability using time-informative gene sets

Quantification of distinguishability of experimental conditions using the indicated time-informative gene sets in Figure 5.1d. Included are so-called confusion matrices resulting from SVMs applied to held-out data (See Methods section ‘Quantification of separability of experimental conditions using gene set scores’), showing the fraction of assigned classes (vertical axes) for cells from each of the indicated experimental conditions (horizontal axes). For a perfect classifier, the diagonals of these matrices are 1 and all off-diagonal entries are 0.

a Cells were stimulated with IFN- γ (100 ng/ml) *in vitro* for the indicated durations. Data depict separability of exposure durations solely using the IFN- γ late gene set score.

b As in a, but now including the IFN- γ plateau gene set score as an additional explanatory variable, resulting in a modest increase in predictive performance (classifier AUROC of .89 from .87).

c As in a, but now using cells stimulated with culture medium (CM) from T cell - tumor cell co-cultures for the indicated durations.

-
- d** As in c, but including the IFN- γ plateau gene set, for a modest increase in predictive performance.
- e** As in c, but employing the TNF- α early gene set.
- f** As in e, but additionally including the TNF- α late gene set for a modest increase in predictive performance.

DISCUSSION

In this thesis, I present a number of studies of cytotoxic T cells in the tumor microenvironment. This work demonstrates that for a multitude of tumors, the landscape of immunologically foreign neoantigens is substantial (**Chapter 2**), allowing for the application of immunotherapy beyond its poster child, melanoma. Following up on this, we evaluated T cell checkpoint blockade therapy in triple negative breast cancer and identified early molecular markers of clinical response (**Chapter 3**), further supporting the notion that T cell attack of varying tumor types is realistic. Moving over to more fundamental research projects on how T cell activity sculpts the tumor microenvironment, we first demonstrated that when T cells encounter cells that present the 'right' (cognate) antigen in the tumor microenvironment, they release IFN- γ that reaches (far) beyond the antigen-presenting cell, whereas TNF- α is strongly localized (**Chapter 5**). We also tested for the loss of neoantigenic mutations as a consequence of T cell pressure in treatment-naïve tumors, but found no evidence for it (**Chapter 4**). A high level summary of these studies would be that T cell recognition of cancer is expected to occur for many types of cancer (**Chapter 2**), can be boosted using combinations of T cell checkpoint blockade and other therapies in TNBC (**Chapter 3**) and that its repercussions reach far and wide in the tumor microenvironment (**Chapter 5**) but do not occur in all matters one would have expected them to (**Chapter 4**). To conclude my thesis, I want to delve deeper into algorithmic requirements to enable highly multiplexed transcriptome based stimulus inference, with the aim of enabling the simultaneous study of many (T cell-secreted) cytokines in the tumor microenvironment.

Cytokines mediate communication between cells of the immune system and other cell types and vice versa, and thereby form key regulators of the immune response. Thus, our understanding of

overall tumor biology and immunology would benefit from the ability to study the dissemination of cytokines and other stimuli with high spatiotemporal resolution. While previous work has shed light on the spatial effects of single or a small number of cytokines simultaneously, a lack of experimental tools for multiplexed investigation has impeded a comprehensive understanding of the spatiotemporal spreading of the many cytokines that collectively mediate and orchestrate (tumor) immunity.

As cytokines effectuate transcriptional changes in the cells that encounter them, a promising way to study their *functional* dissemination (i.e., the reach of their gene regulatory capacity) is to infer it from transcriptional read-outs. This leverages the reporting capability that nature has already granted us, and frees us from having to engineer in reporters ourselves. Given the large diversity in response kinetics among such endogenous reporter genes, we might not be limited to only inferring the nature of the encountered stimuli. Possibly, we could also infer the duration and/or concentration of exposure. Progress in the study of cell communication in general and cytokine dissemination in particular will likely benefit from having available a toolkit for stimulus inference of such nuance.

Here, I will first delve deeper into considerations regarding application of transcription based stimulus inference, that we did not fully explore in **Chapter 5**. Next, I'll discuss ideas for algorithms to deal with the obstacles in this approach. By summarizing the understanding I have built up around this topic, I hope to provide guidance to those interested in building upon our work.

I will start with an overview of the procedure of applying transcriptional/RNA-based stimulus inference (Figure 6.1). First, one would decide whether a specific cell type(s) will serve as *reporter(s)* (**Chapter 5**) or whether a mixture of cell types will be used¹. The reporting cell type could be engrafted in a model system, typically a recipient mouse, but for certain questions an organoid cell culture could also suffice. Alternatively, endogenous cells can also be used, obviating the need for an engraftment step. Next, the experimental model would be exposed to a manipulation of interest (e.g. activation of T cells or other immune cells). Then, the reporter cells would be harvested and can potentially be separated from non-reporter cells using e.g., flow cytometry assisted cell sorting (FACS), which can be facilitated by transducing the reporter cells with a fluorescent reporter before engrafting them in the model system. Cells would then be extracted and sequenced using (single cell) RNA-seq protocols, thereby yielding the *query* data set on which stimulus inference will subsequently be performed. In parallel, one would have to train models of gene expression in the reporting cell type, and possibly also cell viability, in response to cytokines and other stimuli of interest. This can be done using a compendium of *reference* RNA-seq samples to learn from, ideally of the exact same cell type as that of the reporter cell type as different cell types can differ tremendously in their transcriptional response to cytokines¹. To increase power and reduce cost, this reference dataset may be obtained using bulk RNA-seq and is composed of samples that were exposed to the various stimuli of interest. Finally, the analyst would infer the stimulus exposure of the query (single cell RNA-seq) expression data using the cytokine response models trained on the (bulk RNA-seq) reference dataset.

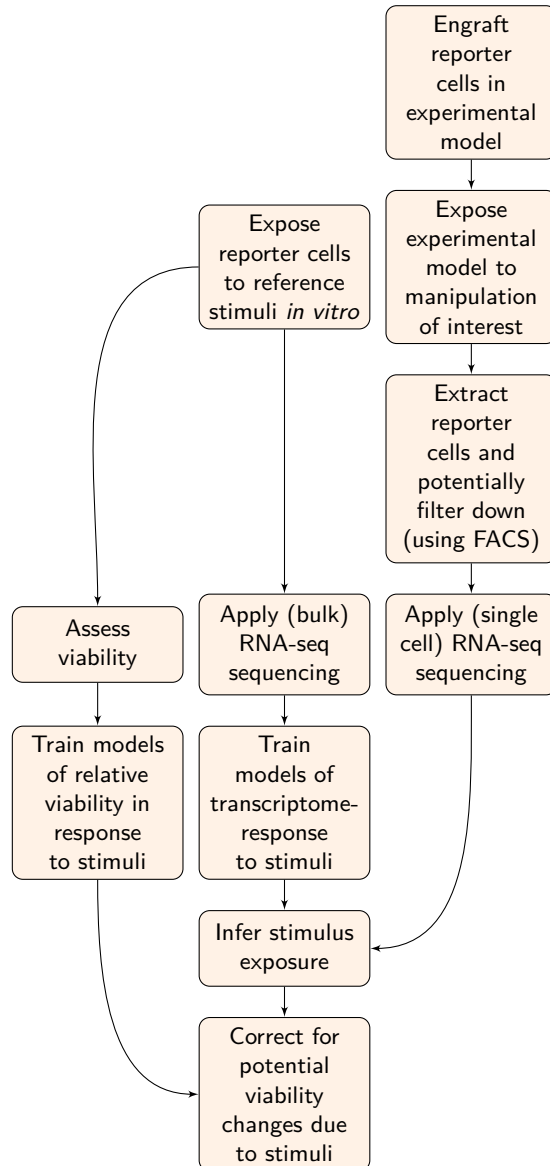


Figure 6.1: An exemplary general process overview for stimulus inference using transcriptomic read-outs

This approach has clear attraction and potential, but there are also a number of potential obstacles to be mindful of in applying it. I will first detail these considerations and then discuss analytical approaches to stimulus inference.

- (i) overlap in reporter genes between studied and also ignored stimuli, complicating discrimination between stimuli.
- (ii) (differing) sources of confounding variation in the query and reference data.
- (iii) modulation of the reporter cell pool's size, by e.g. inducing differentiation into other cell types, cell state transitions, reduced viability or cell death.

(i) Overlap in reporter genes between studied signals

A conceptually simple approach for stimulus inference from (single cell) transcriptome data, is to identify so-called *mono-responsive* genes, which show convincing response to only one of the studied stimuli and minimal to no response to any other stimuli, and sum them for the query transcriptomes. This *gene set* approach should work especially well if the studied cytokines show a high degree of uniqueness, i.e., non-overlap, in terms of their reporter genes. However, the signal transduction pathways that couple stimuli sensing receptors to gene regulatory mechanisms frequently typically display a high degree of crosstalk, causing reporter genes to be shared between stimuli. This can limit the applicability of this approach in two ways. First, the reporter genes may overlap with those of other stimuli in the *in vivo* system of application, at risk of causing false positive inferences if this is not properly accounted for in the experimental design. In **Chapter 5**, the unstimulated (control) samples we collected in addition to the T cell-exposed (test) samples allowed us to establish the baseline level of reporter gene expression signal in absence of the experimental manipulation. Second, overlap between the stimuli of interest will limit our capability of differentiating between them. With a growing number of cytokines or other stimuli, unique reporter genes for the cytokines of interest will become increasingly rare. Supporting this notion, a high degree of overlap in transcriptional response to various cytokines has been found using a large aggregate of publicly available transcriptome datasets². Similarly, related cytokines were shown to overlap in terms of activated genes in single cell data¹. For the murine system we studied in the last part of Chapter 5, mono-responsive genes were scarce to absent. Here, all genes that strongly respond to TGF- β turned out to negatively respond to IFN- γ . Clearly, analytical methodology to deal with this convolvement of factors is a first requirement to studying large numbers of stimuli, especially when these stimuli display a large degree of reporter overlap or when one aims to discriminate between different concentrations and/or durations of the stimuli. In addition to the challenges reporter sharing presents, one clear advantage is that absence of expression of such clusters allows for the inference of absence of not just one but all overlapping stimuli.

(ii) Interference of the transcriptome by confounding sources of variation

A second complication is that the reporter genes, selected for their stimulus reporting behavior, can be subject to modulation by confounding technical factors and/or other biological processes.

A first technical problem may stem from differences in measurement (platform) technology between the (unlabelled) query data and the labelled reference data. Considering the large number of stimuli experimentalists will want to screen to generate reference datasets, bulk RNA-seq is attractive because of its cost effectiveness. However, when using these bulk reference data in conjunction with query single cell data, one will have to deal with the ensuing technical variation between the two platforms. Bulk RNA-seq data typically features millions of reads and detects 10s of thousands of genes per sample, typically showing a rather homogeneous distribution of read counts across genes. In contrast, (droplet-based) single cell RNA-seq only detects a few thousand genes per single cell, with count distributions that are strongly dominated by a relatively small set of genes. When not properly accounted for, the many zeros in the single-cell data would result in low overall similarity to the bulk RNA-seq data, limiting the utility of query-to-reference similarity metrics.

A second complication in relating single cell query to bulk reference data is the potential activity of cellular processes that will be averaged in bulk reference data but clearly not so in the single cell query data. An example of a such a cell-intrinsic process is the cell cycle, which affects thousands of genes in a periodic fashion³. An obvious, experimental remedy for cell-intrinsic sources of variation could, in theory, be to neutralize the interfering process with a drug. However, this risks affecting the transcriptome as well, biasing downstream stimulus inference. A first *analytical* alternative approach is to a priori 'regress out' the confounding process. All popular single cell RNA-seq analysis software packages offer functionality for a simple, statistical correlation-based version of this. However, as the stimuli of interest could also modulate the confounding process (e.g. IFN- γ inhibits cell cycle progression), this could remove much of the informative value pertaining to the stimulus of interest, throwing away the baby with the bath water. A more sensitive approach to offset cell-intrinsic variation could be to apply a deconvolution step in which active processes are identified in a more data-driven, unsupervised manner, allowing for more granular removal of confounding processes. Recent mathematical advances by Karin et al. (2023) have allowed upfront deconvolution of biological processes in (single cell) expression data through spectral analysis of the associated covariance matrix^{4,5}. Rather than relying on pre-identified marker genes of the confounding process to deconvolve the single cell transcriptomes, the authors only assume a 'topology' of variation in gene expression caused by the confounding process⁵. The topology can roughly be understood as the general shape of the path that cells travel in 'state space' as they progress through the confounding process. Mapping the cells to the topology allows for post-hoc identification of genes that align with the topology and filtering of the transcriptome for gene expression that is aligned with the topology. For the specific case of the cell cycle, a circular topology was chosen⁵. Future work could use 'gold standard' single cell datasets, that contain cells exposed to a well-controlled stimulus *in vitro* as reference data, to test whether more complex topologies than the circular one can clean the data

of additional confounding processes and improve accuracy on the downstream stimulus inference task.

Third, stimuli or confounding signals that are restricted to either the reference or to the query settings might (partially) overlap in terms of reporter genes with the stimuli of interest. Probably most prone to this are *in vivo* settings (e.g. the query single cell data in our case), wherein it is likely that many biological processes and cell types are active that are absent from the well-controlled *in vitro* reference conditions of a purified cell type. The possibility of such confounding warrants the design of experiments to include unexposed, control, conditions to any 'test' condition. In **Chapter 5** we included *in vivo* controls without T cells to the *in vivo* test conditions (i.e., with T cells) for this reason. As we'll see below, an analytical approach to remedy this would be a form of data integration that selectively removes covariation that is unique to any of the to be integrated datasets.

For the sake of completeness, I lastly also mention that confounding variation may stem from "mundane" factors like varying batches of chemicals, room temperatures, and different people executing the work.

(iii) Modulation of the reporter cell pool's size

Cytokines and other stimuli of interest may cause (reporter) cells to differentiate into other cell types or to attain differing cell states. They may also directly affect the number of the reporter cells by increasing or decreasing proliferation and even have cytotoxic effects. Unless the reporter cell pool is defined in a wide (i.e., unspecific) enough fashion (and e.g. cell state transitions cannot cause reporter cells to go 'out of scope'), such modulation would bias direct and unadjusted stimulus abundance estimates. The solution will be to explicitly model the effect of such stimuli on the cell pool, allowing for post-hoc adjustment of the direct abundance estimates.

A first example of a stimulus that can modulate the reporter cell pool size is the cytokine IL-2, which specifically induces T cells (more specifically: cells with the IL-2 receptor) to proliferate⁶. One can then immediately see that a direct read-out of the number T cells that *appeared* to have experienced IL-2 cells would not directly reflect the number that have actually done so - the reporter T cells may have simply inherited IL-2-reporting transcripts from their ancestors. Stimuli may also decrease the cell pool size. Among the cytokines we studied in **Chapter 5**, we found IFN- γ to cause reduced proliferation and be cytotoxic to OVCAR5 cells in particular. Dying cells lose their cell membrane integrity and thereby lose their compatibility with FACS-based protocols. As such, if FACS-based enrichment of reporter cells is part of the experimental setup and cytotoxicity remains unaccounted for, the number of cells exposed to cytotoxic stimuli will appear lower than it actually is.

However, increased proliferation or even cytotoxicity (if not complete), does not have to be a show-stopping problem. To attain unbiased estimates of the number of stimulus-experienced cells, one can correct for the expected change in cell numbers due to a given stimulus. For this, one could

use a parametric model of cell growth or death (a so-called viability model), as frequently employed in drug sensitivity screens⁷. If the analyst is additionally willing to assume that the stimulus level remained constant up until the moment of measurement, correcting for cell pool size modulation effects is trivial. First, the viability model would describe viability as a function of stimulus concentration and/or duration with a monotonic, smooth function. Here, viability values larger than 1 would indicate relative cell growth, whereas values smaller than 1 would indicate decline. Next, the number of cells that have experienced the stimulus can be recovered from the direct (biased) estimates. For this, one would multiply the direct estimates by the inverse of the expected change in viability under the stimulus. This concept can be extended to combinations of stimuli as well, given sufficient laboratory resources to combinatorially screen their effects on viability. Alternatively, if one is additionally willing to assume that the cytotoxic effects are not synergistic but rather additive, the cytotoxicity of combinations of stimuli can be modelled by multiplying those of the individual stimuli.

There may be instances in which reference profiles of multiple concentrations and/or exposure durations per stimulus are available and the analyst aims to discriminate and interpolate between these levels. Viability-correction can also be done for such 'continuous' rather than 'discrete' stimulus exposure. Let \mathbf{s} be a vector with entries denoting the concentrations and exposure durations, along with other relevant indicators, of the various stimuli under study. Analogous to the 'discrete' case described above, the computation of summary statistics (e.g., the fraction of cells that have experienced stimulus \mathbf{s} for at least 12 hours) would be done in a weighted rather than unweighted fashion. Here, we would use an integral of the form $\int_D c(\mathbf{s})/v(\mathbf{s})d\mathbf{s}$, wherein D is the domain of interest (e.g., the set of all \mathbf{s} for which IFN- γ -exposure ≥ 12), $c(\mathbf{s})$ is the fraction of cells directly inferred to have experienced stimulus (or the combination of stimuli) \mathbf{s} and $v(\mathbf{s})$ is the estimated viability with this stimulus. A caveat here is that estimation accuracy will decrease with increased cytotoxicity (or extreme degrees of cell growth), and by extension is not feasible for (combinations of) stimuli that fully wipe out the reporter cells.

Regression based stimulus inference

Mindful of the obstacles to transcriptome-based stimulus inference we just considered, how do we choose a stimulus inference methodology? In **Chapter 5**, we actually applied two different ones. The first 'geneset' approach is methodologically simple but sufficed for the human OVCAR5 model, in which mono-reporters were readily available and the experimental design helped us to further disambiguate cytokines. However, for the murine NMM cell line we studied in the same chapter, mono-reporters were virtually absent, necessitating a more sensitive approach. Such an approach should maximally leverage the differential magnitudes with which reporter genes are modulated between stimuli, or different concentrations and durations of one stimulus. Regression analysis, in which gene expression levels are quantitatively compared between query and reference transcriptomes, could be the ideal foundation for this. In theory, it can leverage expressional nuance to dis-

criminate between related stimuli, or even between different exposure durations/concentrations of the same stimulus. Here, a query transcriptome (e.g., a single cell or aggregate of multiple similar single cells) is modelled as a weighted sum of expression profiles from the reference dataset. The weights assigned to each of the reference profiles are then interpreted as proportional to the similarities of these reference profiles to the query transcriptome.

In using regression analysis to relate query (single cell) transcriptomes to (bulk) references, the resulting regression coefficients are sensitive to the way the input data are batch-normalized and integrated. In **Chapter 5** we employed pseudo-bulkification (i.e., aggregating transcriptionally similar cells to 'meta'-cells, Figure 6.2.2) to first make the single cell data more similar to the bulk reference data. We next employed a computational trick to work around the requirement of (near-)perfect comparability. Specifically, we noticed that the presence of a stimulus can not just be shown by high regression coefficients to the associated reference sample, but can also be inferred from obtaining an increased *reconstruction error* when omitting the corresponding reference sample from the analysis. Comparing the 'full model' error, obtained using the full set of reference samples, by a 'partial' one, obtained using only a subset of the reference profiles (i.e., by assessing the relative error of these two models), allowed us to assess the importance of the left out reference samples and the potentially unique gene expression patterns they contained that were not recoverable from any of the remaining samples. As this relative reconstruction error can be computed for individual query transcriptomes (relating to single cells or single cell neighborhoods), the relevance of individual or groups of reference samples to individual query transcriptomes could be quantified. In the absolute sense, the reconstruction error using the full reference set was quite variable between neighborhoods, probably due to a host of possible reasons (as discussed above). However, assessing the relative reconstruction error inherently corrects for this variability. Such a reconstruction error is more frequently used in regression analysis, typically to assess feature importance. However, In the context of (bulk) RNA-seq deconvolution and stimulus inference, this was to the best of my knowledge a novelty.

With all data being derived of one particular cell line and acquired in one single (our) laboratory, the homogeneous nature of the data in **Chapter 5** permitted us to omit stringent batch effect correction. However, using the relative reconstruction error does not fully obviate the need for sensitive correction of confounding variation. A computational methodology that can appropriately deal with data from heterogeneous sources benefits from being able to draw upon a much broader set of data sets and sources. With this in place, one could compile a compendium of cytokine stimulated transcriptomes from a range of studies², maximally leveraging past efforts of the scientific community, and arrive at a more diverse and powerful predictor of cytokine exposure. To leverage such heterogeneous data, one needs to deal with confounding variation 'upfront' and preprocess the data in a manner that filters out (biological or technical) confounding variation. How do we arrive at this point?

A first obstacle is the difference in count distributions between bulk and single cell RNA-seq data (see above), that precludes direct use of preprocessing solutions developed specifically for single cell

RNA-seq data. This is probably why I found that existing single cell RNA-seq integration methodologies, which essentially rely on library size normalization and Euclidean distance-based similarity metrics to relate transcriptomes, did not perform well when applied to the OVCAR5 data from Chapter 5 (data not shown). However, such algorithms did achieve sensible integrations for their intended domain of application, i.e., data sets that were exclusively of single cell nature. If one has access to both bulk and single cell RNA-seq data of the same cell lines and stimuli - like we did - then one viable approach may be to somehow learn how to simulate single cell transcriptomes from bulk RNA-seq samples, such that these simulated cells can then be related to the query single cells. Specifically, one could use a modification of the variational auto-encoder (VAE) for this, which would disentangle the various active transcriptional programs from each other. The VAE would consist of two (deep) neural networks: one encoder network which reduces the dimensionality of bulk reference gene expression to a much smaller number of latent factors, and a second decoder network which reconstructs single cell expression profiles from the latent factors. Each latent factor should represent a (biological) source of variation and captures combinations of correlated non-linear patterns in the data, akin to the principal components in a (linear) principal component analysis. Input to the probabilistic VAE encoder would be the bulk expression data, along with stimulus exposure and batch-related metadata, and the output of the VAE decoder would be the associated single cell transcriptomes. To ensure biological relevance of the identified latent factors, sparsity constraints could be imposed on the VAE's component models, limiting the number of active latent factors per sample. Another approach towards this goal would be to pre-train the VAE on single cell data exclusively. This is less ambitious as it doesn't require the encoder network to additionally learn which processes tend to co-occur in single cells, like it would have to when its input is bulk data. Then, especially the encoder network could be adapted to the task of using bulk data as its source (i.e., an application of transfer learning), or an additional encoder network could be placed in front of the encoder network. Having the VAE in place, one could next simulate tons of single cells for each of the reference samples, including the ones for which no matching ground truth single cells are available. To then finally do stimulus inference for the query *in vivo* data, the frequently used and robust mutual nearest neighbor algorithm⁸ could be used to integrate them with the labelled, simulated cells in an unsupervised manner.

Another approach to leveraging data from different labs or lab technologies is to directly relate query and reference transcriptomes in a manner that is robust to distributional differences and confounding variation. In collaboration with Soufiane Mourragui, I developed an approach to integrate data such that it is insensitive to differences in count distributions and that resulted in a good balance between inferential sensitivity and robustness (Data Integration inSensitive To distributional iN-balanCes, DISTINcT, Figure 6.2. It involves relating bulk RNA-seq and single cell RNA-seq data in a higher-order feature space induced by the *Mallow's kernel*⁹ combined with Mourragui's domain adaptation algorithm TRANSACT¹⁰. On our own *in vitro* data, it remained completely insensitive to the distributional differences between the bulk and single cell RNA-seq data, while nearly perfectly inferring IFN- γ stimulus exposure duration (Figure 6).

DISTINCT

Data Integration in Sensitive To distributional iNbalances

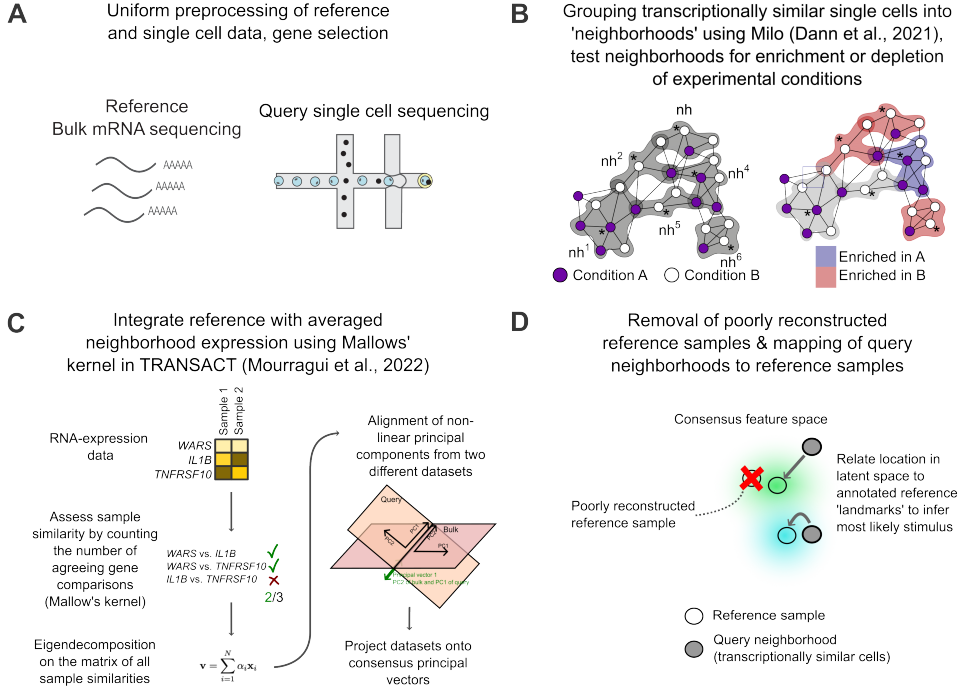
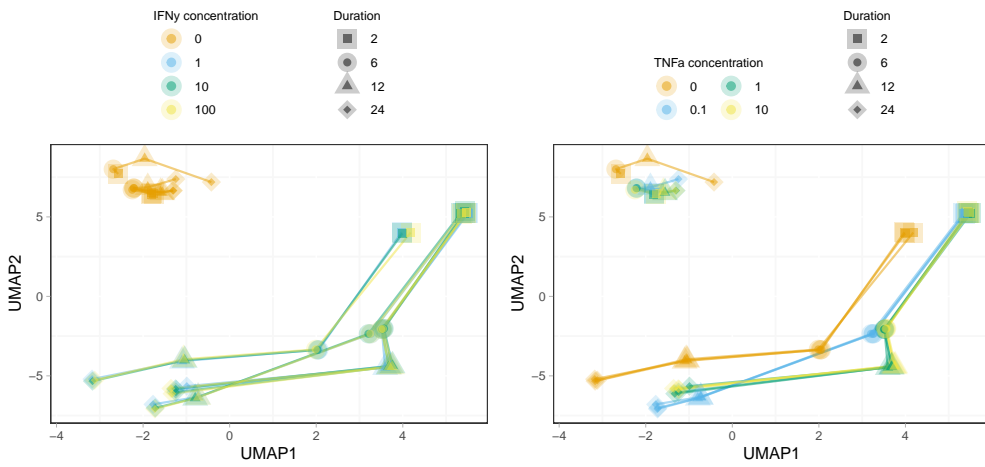


Figure 6.2: The Mallows' kernel can be seen as a mathematical function to compute sample similarities, using gene ordering (i.e., all pairwise comparisons between genes) instead of absolute expression levels to relate sample to each other. This makes it robust against the distributional differences between bulk and single cell sequencing samples. From the pairwise sample similarities within any of the individual datasets, one can derive a latent, lower dimensional representation using kernel principal component analysis, a generalization of 'standard' PCA. Oversimplifying, the non-linear principal components from the kernel PCA (with the Mallows' kernel) group comparisons of genes that tend to co-occur. TRANSACT builds on kernel PCA by matching sets of non-linear principal components of the two datasets with each other and interpolating between them to compute so-called consensus features (C). The consensus features capture sources of variation (e.g., gene expression programs) within each dataset that are shared between the two datasets, discarding variation that is unique to any one of the two individual datasets. This makes it robust to, among other confounders, cell-intrinsic sources of covariation (e.g. the cell cycle). It finishes by projecting the reference and query datasets onto the consensus space, allowing direct comparison between the reference and query samples. Proximity in the consensus space can be interpreted as an indication of similar cytokine exposure (D). To further increase inferential strength and speed up computation, one can optionally also a priori aggregate single, transcriptionally similar cells into local averages and thereby denoise RNA expression values, using cell graph based aggregation^{11,12} (B).

Bulk reference



Single-cell query

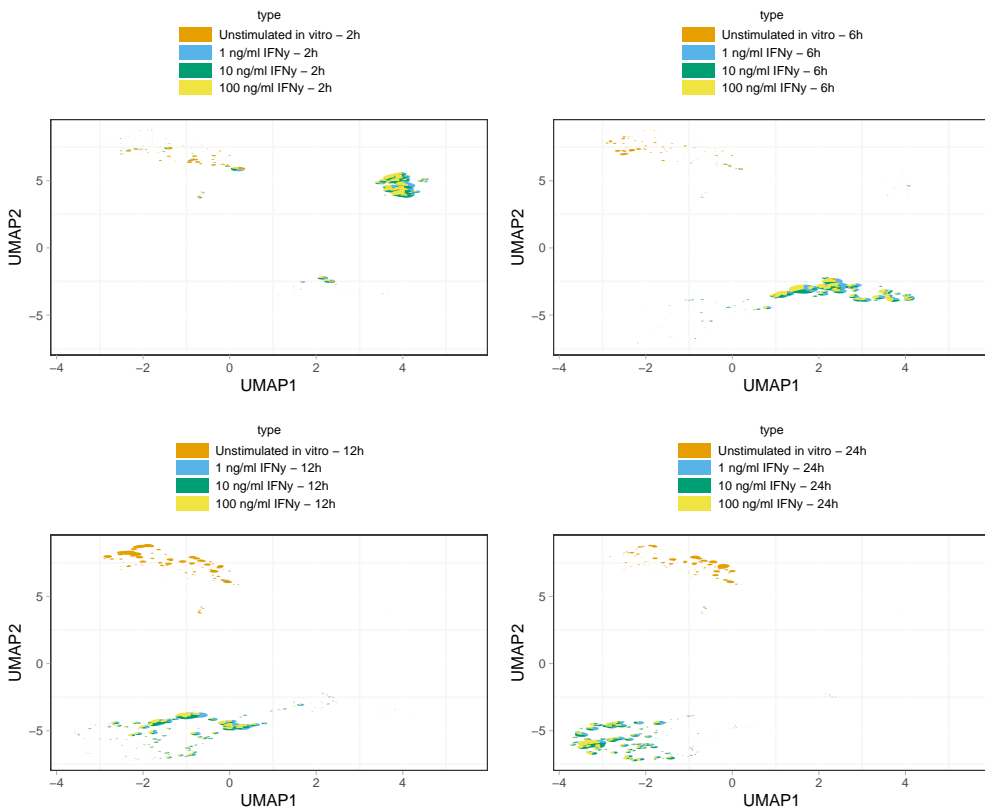


Figure 6.3: Seamless integration of reference bulk and *in vitro* single cell query data sets with DISTINcT. All panels show a UMAP-transformation of samples projected onto the consensus features resulting from TRANSACT-integration in combination with the Mallow’s kernel. The top 2 panels show the bulk RNA-seq reference dataset, once colored by IFN- γ -concentration (top row, left) and once by TNF- α -concentration (top row, right). The lines connect samples that were stimulated with the same stimuli and concentrations. Since we controlled the stimulus exposure of these cells experimentally *in vitro*, we can use them for benchmarking purposes, along with unstimulated controls. The bottom 4 panels show the *in vitro* IFN- γ -titration single cell query dataset, with one figure panel per exposure duration. Each circle (pie-chart) in these panels represents one ‘neighborhood’ of cells, i.e., a variably sized group of transcriptionally similar cells, that is color coded by the exposure concentration of IFN- γ . This single cell dataset shown here consists of all 16 combinations of 4 concentration of IFN- γ (0, 1, 10 and 100 ng/ml) and 4 exposure durations (2, 6, 12 and 24h). The reference reference dataset (top panels) covers these same stimuli but additionally also contains 3 TNF- α concentrations, as well as the combination of IFN- γ and TNF- α . The integrated data permit clear inference of IFN- γ exposure time, but not concentration, for the bulk reference and single cell query data. This can be seen by the clear separation of exposure durations for both the reference and query data. For instance, the first panel (middle row, left) shows the coordinates of cells that were stimulated for 2 hours. Here, unstimulated cells (orange) gravitate towards the top left in UMAP space, where the unstimulated reference samples are also positioned. In contrast, all IFN- γ -stimulated cells are located on the top right, exactly where the bulk reference data of this duration are also positioned. For the other exposure durations, the single cell query and bulk reference data are also well aligned. This shows that the Mallow’s kernel represents the input data in an informative manner, capable of discerning different exposure durations of IFN- γ to each other, whereas different concentrations of IFN- γ at any of the evaluated exposure durations are too similar in order to be reliably distinguished. However, this latter unseparability was already present in the untransformed, original data (data not shown). This indicates that gene expression just cannot discriminate between these IFN- γ concentrations. Also, notice how the TNF- α -stimulated cells in the reference are positioned very close to the unstimulated samples. Due to the absence of TNF- α -stimulation in the single-cell query data, TRANSACT mostly discards TNF- α -related variation in the bulk RNA-seq data (*‘variational collapse’*).

Currently still open is the question on how to deal with *‘variational collapse’* in DISTINcT. DISTINcT finds consensus features using shared variability, and variability that is unique to any one of the datasets will hence - by design - be discarded. This has the consequence that samples that have been stimulated in a manner that uniquely occurs in just one of the two datasets to be integrated will appear to not or barely have sensed this stimulus in the integrated data representation. Variational collapse is observable in Figure 6, wherein the reference samples stimulated with both IFN- γ and TNF- α appear eerily close to the reference samples that were exposed to IFN- γ alone. This is due to the absence of TNF- α -stimulated samples in the query dataset, such that all gene comparisons that are informative in that regard are discarded in the consensus space. One way to (automatically) diagnose this problem would be to assess how accurately the consensus space representation resembles the original input representation (i.e., the reconstruction error), but the mathematics to support this operation have not been developed yet. To get a proxy for the reconstruction error, the relative distance between samples in the original and consensus representations can be considered. For instance, to quantify the reconstructability of sample j , one could consider its distance (i.e., the l^2 -norm) to an unstimulated ‘anchor’ sample in the consensus representation, and divide that by the same sample ratio in the original/untransformed representation. Samples for which this reconstructability value lies near 2 could be deemed as equally well represented in both the original and the

integrated representations. In contrast, values closer to 0 would indicate the imprint of a stimulus or process that was not represented in the consensus space, presumably because it wasn't present in the other dataset. To offset such variational collapse, the last query to reference mapping step of the algorithm would weight reference samples based on their reconstruction error, prioritizing samples with high reconstructability, or fully remove (reference) samples for which the reconstruction does not meet some predefined criterion (Figure 6.2.D).

Additional inferential value from alternative splicing information

Isoform specific mRNA quantification, along with large numbers of samples as are typically screened in single cell RNA-seq experiments, allow for estimation of so called RNA velocity¹³. RNA velocity is a time derivative of gene expression which can be used to order cells along temporal and cell differential axes. It might also offer a rich layer of information regarding the timing of cytokine exposure, complementing that of the 'snapshot' RNA abundance-derived estimates that are discussed above and in **Chapter 5**.

In exploratory analyses with our bulk RNA-seq data, which was not acquired using unique molecular identifiers (UMIs) as is typically done in single cell RNA-seq, we obtained noisy estimates of RNA velocity that did not offer informative value in addition to the readily available splicing-agnostic gene expression estimates. However, in the single cell RNA-seq data we acquired using the 10x protocol and with each identified molecule tagged with a UMI, RNA velocity should be much more robustly inferrable. Even though the per sample sequencing depth in single cell RNA-seq is much lower, the large number of samples in conjunction with population based estimates¹³ in these datasets offer an opportunity towards increasing the signal-to-noise ratio.

To then properly incorporate RNA velocity in the frameworks that are discussed above, one could set up a reference compendium with single cell RNA-seq. Compiling a reference dataset of single cell data would be more costly but would also give more insight in terms of cell-to-cell response heterogeneity and extrinsic sources of variation. An exciting newer technology to use is VASA-seq, which unlike more traditional single cell-seq protocols is not biased towards the 3' ends of RNA molecules and allows for more accurate RNA velocity estimates¹⁴. VASA-seq is also applicable to bulk RNA-seq sequencing (personal communication with Soufiane Mourragui). I expect RNA velocity features to especially become useful when deconvoluting transcriptomes affected by many cytokines of overlapping transcriptional signatures. On a related note, multimodal, rather than unimodal, data, for instance additionally employing CITE-Seq¹⁵ with antibodies directed against cell surface proteins relevant to the stimulus in question, could also allow for further increases in inferential accuracy.

Outlook of transcriptome-based cytokine inference

The work my colleagues and I have presented in Chapter 5 forms a useful step in our understanding of cytokine mediated communication between T cells and the tumor microenvironment. I foresee that in the decades to come, the field will continue to generate research tools to study signaling with increasingly high throughput. This increasing throughput is likely to be valuable, as the complex underpinnings of the tumor microenvironment and other biological systems can only be fully elucidated by assessing these dynamic entities in parallel. The increasingly widespread use of spatial single RNA-seq, and its ability to study cells in their physiological context¹⁶, will unlock another crucial layer of information with regard to cytokine dissemination and amplification. In this, deconvolution of the effects of the many cytokines that may be at play will be especially important. As transcriptome based inference has high discriminatory potential and does not require upfront modification of the reporter cell, it's bound to continue to be a potent avenue towards studying stimulus dissemination and cellular crosstalk in biological systems.

References

1. Cui, A. *et al.* Dictionary of Immune Responses to Cytokines at Single-Cell Resolution. *Nature*, 1–8. ISSN: 1476-4687. (2023) (Dec. 2023).
2. Jiang, P. *et al.* Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nature Medicine* **24**, 1550–1558. ISSN: 1546-170X (Oct. 2018).
3. Dominguez, D. *et al.* A High-Resolution Transcriptome Map of Cell Cycle Reveals Novel Connections between Periodic Genes and Cancer. *Cell Research* **26**, 946–962. ISSN: 1748-7838. (2024) (Aug. 2016).
4. Nitzan, M. & Brenner, M. P. Revealing Lineage-Related Signals in Single-Cell Gene Expression Using Random Matrix Theory. *Proceedings of the National Academy of Sciences* **118**, e1913931118. (2024) (Mar. 2021).
5. Karin, J., Bornfeld, Y. & Nitzan, M. scPrisma Infers, Filters and Enhances Topological Signals in Single-Cell Data Using Spectral Template Matching. *Nature Biotechnology* **41**, 1645–1654. ISSN: 1546-1696. (2024) (Nov. 2023).
6. Ross, S. H. & Cantrell, D. A. Signaling and Function of Interleukin-2 in T Lymphocytes. *Annual review of immunology* **36**, 411–433. ISSN: 0732-0582. (2024) (Apr. 2018).
7. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*. ISSN: 10974172 (2016).
8. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch Effects in Single-Cell RNA-sequencing Data Are Corrected by Matching Mutual Nearest Neighbors. *Nature Biotechnology* **36**, 421–427. ISSN: 1546-1696. (2024) (May 2018).
9. Jiao, Y. & Vert, J.-P. The Kendall and Mallows Kernels for Permutations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**, 1755–1769. ISSN: 0162-8828, 2160-9292. (2023) (July 2018).
10. Mourragui, S. M. *et al.* Predicting Patient Response with Models Trained on Cell Lines and Patient-Derived Xenografts by Nonlinear Transfer Learning. *Proceedings of the National Academy of Sciences of the United States of America* **118**. ISSN: 10916490. (2022) (Dec. 2021).
11. Baran, Y. *et al.* MetaCell: Analysis of Single-Cell RNA-seq Data Using K-nn Graph Partitions. *Genome Biology* **20**, 206. ISSN: 1474-760X. (2023) (Oct. 2019).
12. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential Abundance Testing on Single-Cell Data Using k-Nearest Neighbor Graphs. *Nature Biotechnology* **40**, 245–253. ISSN: 1087-0156, 1546-1696. (2023) (Feb. 2022).
13. La Manno, G. *et al.* RNA Velocity of Single Cells. *Nature* **560**, 494–498. ISSN: 1476-4687. (2024) (Aug. 2018).

-
14. Salmen, F. *et al.* High-Throughput Total RNA Sequencing in Single Cells Using VASA-seq. *Nature Biotechnology* **40**, 1780–1793. issn: 1546-1696. (2024) (Dec. 2022).
 15. Stoeckius, M. *et al.* Simultaneous Epitope and Transcriptome Measurement in Single Cells. *Nature Methods* **14**, 865–868. issn: 1548-7105. (2024) (Sept. 2017).
 16. Schäbitz, A. *et al.* Spatial Transcriptomics Landscape of Lesions from Non-Communicable Inflammatory Skin Diseases. *Nature Communications* **13**, 7729. issn: 2041-1723. (2024) (Dec. 2022).

SUMMARY

This thesis delves into the ways that T cells interact with the tumor microenvironment, the types of tumors that they can be expected to see as immunologically foreign, whether such antigen-recognition detectably leads to the specific destruction of antigen-carrying cells and how to enhance T cell activity to treat cancer patients, specifically those with triple negative breast cancer.

The span of the questions that are addressed in this thesis is perhaps rather broad, considering that present-day PhD theses are supposed to report deeply on an increasingly more narrow topic. What ties the individual projects together, beyond the theme of T cells in the cancer microenvironment, is the need for computational approaches to answer them - my primary area of interest. During my time at the Netherlands Cancer Institute, I have worked on projects that had direct *translational* potential, questions that could directly inform care for patients with cancer. I have also been interested in improving our fundamental understanding of T cell biology, as such insights can eventually, but not necessarily within the span of a PhD, have an impact on clinical practice. The history of science demonstrates that the greatest technological achievements could not have originated in absence of a prior fundamental understanding to fuel them. On a more personal note, exploring nature for the sake of it is in my view oftentimes also more intellectually stimulating.

Taking a step back on cancer, we see a large diversity in the cells that make up a tumor. Cancer cells arise by the accumulation of DNA damage in our own, healthy cells. But the tumors that such cells form are not solely composed of mutated, cancer cells. Many unmutated host cells will also be recruited to contribute to the growing tumor mass. These cells may - for example - provide structural support, or facilitate the perfusion of blood flow through the tumor, in turn important for the dis-

tribution of nutrients and oxygen. Various cell types of the *immune system* will also migrate to the nascent tumor mass. The immune system has classically been understood to clear pathogens, such as bacteria and viruses, but a clear role in the control of cancer, and oftentimes also its facilitation, is now a part of our immunological understanding as well. Cytotoxic T cells, those positive for the cell surface marker CD8⁺ and capable of killing cells that they can form interactions with, can respond to antigens that directly results from DNA damage. These *neoantigens* are especially attractive from a therapeutic perspective as their expression is restricted to tumor cells. Hence, boosting neoantigen-reactive T cells should result in high *on target* activity, and be free of the *off-target* activity that more traditional treatment modalities (surgical, chemotherapeutic, and radiation therapies) are limited by.

The first part of this thesis is translational, focusing on questions that have the potential to directly affect the clinical care for patients with cancer. We now know that neoantigen-reactive T cells are frequently integral to the clinical efficacy of T cell boosting cancer immunotherapies. Taking a helicopter view in **Chapter 2**, we mapped the number of neoantigens and its diversity across human malignancies. Encouragingly, we found ~50% of assessed cancer samples to be sufficiently rich in predicted (neo)antigens for at least minimal T cell recognition, suggesting widespread applicability of T cell engaging therapies in cancer treatment. This means that a first critical requirement for T cell-mediated tumor control (i.e., 'immunological foreignness') is met. However, it does not imply that all these cancers respond to treatment. Specifically, tumors can throw up ingenious barricades and subvert the immune system in order to preserve themselves during immune attack. Hence, successful immune-mediated tumor clearance will require all environmental conditions to be conducive to tumor clearance, besides the primary requirement for a 'recognition point' (i.e., antigen) on the tumor cell surface.

In **Chapter 3**, we study the application of T cell checkpoint blockade in a triple negative breast cancer. Breast cancer as a whole may not immediately have been expected to frequently respond to immunotherapy, as it is generally lowly infiltrated by immune cells. However, *triple negative* breast cancer is rather highly mutated and so should at least have sufficient antigenic diversity to facilitate T cell recognition. The therapeutic class of T cell checkpoint inhibitors aims to stop a blockade that (cancer) cells can throw up to inhibit T cells and tone down their activity. Importantly, this can only be of use if the tumor is infiltratable by T cells in the first place. In order to increase immune infiltration before administration of T cell checkpoint inhibitors, we assessed different *induction* treatments. Using extensive molecular characterization, we could trace how these induction treatments affected the tumor microenvironment. We learned what clinical responses look like on a molecular level - i.e., massive immune cell activity - and, using longitudinally collected samples, could compare induction therapies in terms of their capacity to yield this immunoactive profile. Although we did see differences between induction therapies, the strongest T cell activating effects could be ascribed - not entirely unexpectedly - to the T cell checkpoint inhibition itself. This latter conclusion has since been replicated in many other studies.

The second part of this thesis delves into fundamental T cell biology in the context of cancer. T cells

are known to release a range of different signalling molecules (cytokines) upon their activation by antigen-positive ('visible') cells. In **Chapter 5**, we investigate the effects of T cells on tumor cells they could not directly recognize, in order to measure the range of cytokine spreading. Of these, based on its mode of secretion, TNF- α has been thought to diffuse far beyond the antigen-positive cells, whereas IFN- γ is thought to act in a more localized fashion, i.e., restricted to the target cell that is being engaged. Given the known effects of IFN- γ (upregulation of MHC molecules that are required for T cell recognition, increased expression of T cell inhibitory receptors), this localized action appeared illogical. Perhaps akin to sounding a battle horn, IFN- γ 's stimulation of antigen presentation machinery, but also induction of possible feedback mechanisms in case of profound T cell activity, would appear most beneficial to ready tissues for T cell screening, and not as much to modify the transcriptome of (cancer) cells that are already being screened by T cells. In line with this theoretical argument, we found IFN- γ to have large 'field' effects, whereas TNF- α 's scope is more focal. Additionally, IFN- γ -experienced cells showed reduced TGF- β signaling, in a manner that appeared orthogonal to IFN- γ 's direct effect on TGF- β -responsive genes, indicating a potential shift in (immune) cell activity around IFN- γ -experienced cells.

In **Chapter 4**, we revisit the neoantigen predictions of **Chapter 2** to explore the selective pressure exerted by T cells on developing cancers. Despite robust methodology, this study did not find evidence of neoantigen depletion in treatment-naïve tumors. Several factors could explain this, including limitations of current neoantigen identification capabilities and the possibility that tumors primarily use other mechanisms besides the (genetic) loss of neoantigens to evade the T cell based immune system.

This thesis concludes by discussing future directions for expanding the methodology developed in **Chapter 5**. With further algorithmic innovation, RNA-based signal inference could extend beyond the relatively small range of cytokines we studied in **Chapter 5**. Such methodology could further enhance our understanding of immune interactions in cancer and other immune-related diseases, and thereby inform the development of more effective immunotherapies.

SAMENVATTING

Dit proefschrift onderzoekt hoe T-cellen interacteren met het tumor micromilieu, welke typen tumoren zij kunnen herkennen als immunologisch vreemd, of die antigeenherkenning daadwerkelijk leidt tot specifieke vernietiging van antigeen-dragende cellen, en hoe T-celactiviteit versterkt kan worden om kankerpatiënten te behandelen, in het bijzonder patiënten met triple-negatieve borstkanker.

De reikwijdte van de vragen die in dit proefschrift aan bod komen is wellicht breed, zeker gezien de huidige trend waarin promotieonderzoeken een steeds nauwer onderwerp uitdiepen. Wat de individuele projecten hier met elkaar verbindt, naast het thema van T-cellen in het kankermicromilieu, is de noodzaak van computationele benaderingen om ze te beantwoorden – mijn voornaamste interessegebied. Tijdens mijn tijd bij het Nederlands Kanker Instituut heb ik gewerkt aan projecten met directe *translationele* potentie – vragen die direct de zorg voor kankerpatiënten kunnen beïnvloeden. Daarnaast heb ik mij geïnteresseerd in het verbeteren van ons fundamentele begrip van T-celbiologie, omdat zulke inzichten uiteindelijk, zij het niet noodzakelijk binnen het tijdsbestek van een promotietraject, de klinische praktijk kunnen beïnvloeden. De geschiedenis van de wetenschap laat zien dat grote technologische doorbraken zelden zonder een voorafgaand fundamenteel begrip tot stand zijn gekomen. Daarnaast is het verkennen van de natuur, puur uit nieuwsgierigheid, vaak ook intellectueel het meest stimulerend.

Als we een stap terug doen bij kanker, zien we een grote diversiteit aan cellen die een tumor vormen. Kankercellen ontstaan door ophoping van DNA-schade in onze eigen, gezonde cellen. Maar de tumoren die hieruit ontstaan bestaan niet alleen uit gemuteerde kankercellen. Veel ongemuteerde

cellen van de gastheer worden eveneens gerekruteerd om bij te dragen aan de groeiende tumormassa. Deze cellen kunnen bijvoorbeeld structurele ondersteuning bieden of de doorbloeding van de tumor faciliteren, wat weer belangrijk is voor de distributie van voedingsstoffen en zuurstof. Verschillende celtypen van het *immuunsysteem* zullen zich ook naar de zich ontwikkelende tumor begeven. Het immuunsysteem werd klassiek gezien als middel om pathogenen, zoals bacteriën en virussen, te elimineren, maar tegenwoordig is ook de rol in het controleren – en soms zelfs faciliteren – van kanker een erkend onderdeel van ons immunologisch begrip. Cytotoxische T-cellen, die positief zijn voor het celoppervlakte-eiwit CD8⁺ en in staat zijn cellen waarmee ze interacteren te doden, kunnen reageren op antigenen die direct voortkomen uit DNA-schade. Deze *neoantigenen* zijn therapeutisch gezien bijzonder aantrekkelijk, omdat hun expressie is beperkt tot tumorcellen. Het versterken van T-cellen die gericht zijn tegen zulke neoantigenen zou daarom moeten leiden tot hoge *on-target* activiteit, en dat zonder de *off-target* toxiciteit waar traditionele behandelmodaliteiten (chirurgie, chemotherapie en radiotherapie) vaak door worden beperkt.

Het eerste deel van dit proefschrift is translationeel van aard, gericht op vragen die de klinische zorg voor kankerpatiënten direct kunnen beïnvloeden. We weten inmiddels dat T-cellen gericht tegen neoantigenen vaak cruciaal zijn voor de effectiviteit van kankerimmunotherapieën die T-cellen versterken. In een overzichtsstudie in **Hoofdstuk 2** brachten we het aantal en de diversiteit van neoantigenen in menselijke maligniteiten in kaart. Bemoedigend was dat we in ~50% van de onderzochte monsters voldoende voorspelde (neo)antigenen vonden voor ten minste minimale T-celherkenning, wat wijst op een brede toepasbaarheid van T-celgerichte therapieën voor kankerbehandeling. Dat betekent dat aan een eerste vereiste voor T-cel-gemedieerde tumorcontrole (namelijk ‘immunologische vreemdheid’) is voldaan. Maar dat betekent niet dat al deze kankers ook daadwerkelijk reageren op behandeling. Specifiek kunnen tumoren ingenieuze barrières opwerpen en het immuunsysteem omzeilen om zichzelf te beschermen tegen een aanval. Succesvolle immuun-gemedieerde tumor-eliminatie vereist daarom dat naast het ‘herkenningspunt’ (i.e. antigeen op de tumorcel) ook de rest van de omgevingscondities bijdragen aan tumorcontrole door T-cellen.

In **Hoofdstuk 3** bestuderen we de toepassing van checkpoint-remmers in triple-negatieve borstkanker. Borstkanker als geheel stond niet bekend als gevoelig voor immunotherapie, omdat het doorgaans weinig immuuncelinfiltratie vertoont. Triple-negatieve borstkanker is echter relatief hoog gemutueerd, en zou daarom voldoende diversiteit aan antigenen kunnen bevatten voor T-celherkenning. Checkpointremmers trachten een remmend signaal op te heffen dat (kanker)cellen kunnen gebruiken om T-cellen af te remmen. Cruciaal is dat dit alleen werkt als de tumor überhaupt door T-cellen geïnfilteerd kan worden. Om de infiltratie voorafgaand aan checkpointremming te verhogen, onderzochten we verschillende *inductiebehandelingen*. Via uitgebreide moleculaire karakterisering konden we nagaan hoe deze behandelingen het tumormilieu beïnvloeden. We leerden hoe klinische respons eruitziet op moleculair niveau – i.e. massale immuuncelactiviteit – en konden met longitudinaal verzamelde monsters inductietherapieën vergelijken op hun vermogen om zo’n immuunactief profiel op te wekken. Hoewel we verschillen tussen behandelingen zagen, waren de sterkste effecten op T-celactivatie – niet geheel onverwacht – toe te schrijven aan de checkpointremming zelf. Deze

conclusie is inmiddels in meerdere studies bevestigd.

Het tweede deel van dit proefschrift gaat over fundamentele T-celbiologie in de context van kanker. T-cellen staan bekend om het afgeven van signaalmoleculen (cytokines) wanneer zij worden geactiveerd door antigeen-positieve ('zichtbare') cellen. In **Hoofdstuk 5** onderzoeken we de effecten van T-cellen op tumorcellen die zij niet direct konden herkennen, om de reikwijdte van cytokinespreiding te meten. Van deze cytokines dacht men dat TNF- α , op basis van zijn secretiewijze, zich ver verspreidt vanuit antigeen-positieve cellen, terwijl IFN- γ juist lokaal zou werken, i.e. beperkt tot de cel waarmee interactie is. Gezien de bekende effecten van IFN- γ (zoals verhoogde expressie van MHC-moleculen en T-celremmende receptoren) leek dit lokale effect onlogisch. Het lijkt logischer dat IFN- γ , als een soort strijdkreet, de omgeving voorbereidt op T-celscreening, in plaats van de transcriptie van al herkende kankercellen te beïnvloeden. In lijn met dit theoretische argument vonden we dat IFN- γ brede 'veld'-effecten had, terwijl TNF- α juist meer lokaal werkte. Bovendien lieten cellen die blootgesteld waren aan IFN- γ een verminderde TGF- β -signaaltransductie zien, op een manier die orthogonaal leek aan het directe effect van IFN- γ op TGF- β -responsieve genen, wat wijst op een mogelijke verschuiving in (immuun)celactiviteit rondom door IFN- γ beïnvloede cellen.

In **Hoofdstuk 4** keren we terug naar de neoantigeenvoorspellingen van **Hoofdstuk 2** om de selectiedruk van T-cellen op zich ontwikkelende tumoren te bestuderen. Ondanks robuuste methodologie vonden we geen bewijs voor neoantigeen-depletie in onbehandelde tumoren. Meerdere factoren kunnen dit verklaren, waaronder de huidige beperkingen in het identificeren van neoantigenen, en de mogelijkheid dat tumoren voornamelijk andere mechanismen gebruiken dan het genetisch verliezen van neoantigenen om T-celherkenning te ontwijken.

Het proefschrift sluit af met een blik op de toekomst van de in **Hoofdstuk 5** ontwikkelde methodologie. Met verdere algoritmische innovaties zou RNA-gebaseerde signaalinferentie zich kunnen uitbreiden naar een breder scala van cytokines dan wat in **Hoofdstuk 5** is bestudeerd. Dergelijke methodologie zou ons begrip van immuuninteracties in kanker en andere immuun-gerelateerde ziekten verder kunnen verdiepen, en zo bijdragen aan de ontwikkeling van effectievere immunotherapieën.

ACKNOWLEDGMENTS

Ik heb het zo lang mogelijk uitgesteld, maar nu is het toch echt afgerond. Ik ben velen dankbaar voor deze bijzondere fase in mijn leven.

Ton, mijn idee van wat goed wetenschappelijk onderzoek inhoudt, werd in jouw groep volledig herzien. Ik probeer nog altijd jouw scherpe blik en vermogen om complexe vraagstukken te doorgronden enigszins te evenaren.

Lodewyk, ik bewonder je om je kalmte te midden van de grote diversiteit aan projecten waarin je betrokken bent. Consciëntieus ontleed je nieuwe methodes, onderscheid je zin van onzin en - als het even tegen zat - hielpen je bemoedigende woorden me verder.

Joost, dank voor de introductie tot dit fascinerende vakgebied. Ook van jou leerde ik over methodologie.

Mirjam, ons gezamenlijke project verraste mij. Ik leerde dat samenwerken leuk kan zijn en zelfs vonden we iets interessants tussen de koffietjes, crises en levensbeschouwingen door.

Leonie en Marleen, ik ben heel blij dat bij jullie ben ingestapt in TONIC. Jullie bevologenheid en patiëntbetrokkenheid zorgden ervoor dat ik het maximale uit de beschikbare data wilde halen.

Dank aan iedereen van B3 en later B6. Wie had gedacht dat wetenschappers zo levendig en kleurrijk konden zijn? Feline, Anne, Meike, Kaspar, Lianne, Mirjam — en ook jij, Max: wat was het fijn met jullie in de Schumi 'Band of Brothers': hard op de inhoud, zacht voor elkaars meer excentrieke trekjes. Jos, waar tref ik ooit nog iemand zo onbevangen als jij? Mireille, voor je originaliteit en hart voor

anderen. Leonie, Lorenzo, Ana, Francesca, Daniela, Daisy, Leonie, Lisanne, Kelly, I loved the homely feel of our B6 sardine can (aka our office), even if sometimes I had to retreat into white noise to get some work done. Brenda, for your endless curiosity. Lorenzo, Telma, Marit, Raquel, Leïla, Marcel, Marlous, Marcel, Jules, Riccardo, Camilla, Chong, Joost, Sander, Renate, Wouter, Noor, Rhianne, Pia, Silvia, Paula's 'V' and 'K', Kelly, for making the (early stages of) my PhD so incredibly fun. Krijn, Joris, Steven, Tom, Kong, Thomas, Oscar, David, Julia, Georgi, Sofia, for sparring and some banter at the coffee machine. Huiwen and Lisa, for your help and the opportunity to practice mentoring.

Arno, Roel, Iris, Marja, Ron van de Genomics Core Facility, dank voor jullie werk in software en sequencing!

A huge thanks also to everyone in Lodewyk's group, for fostering such a supportive and engaging community — and for keeping the whiteboards in perpetual flux. I want to highlight Gergana, Kat, Bram, Nanne, Julian, Daniël, Ewald, Tycho, Alberto, Tim, Joe, Marlous, Kathy, Sylvana, Tesa, Alex, Misha, Evert, Duco, Lindsay, Olga, Maksim, Michael, and also Patty and Åsmund. Soufiane, for showing methodological excellence without a trace of hubris. Nanne, and also Marcelo and Rocco, for all of the impromptu music.

Thanks to my (current) CureVac colleagues: Katka, Marlon, Seline, Jeroen, Michael, Wigard, Ronald, Katharina, Hugo, Lisa, Olga, Aroa, Mark, Erik, Renée, Jamie, Wim, your support has strengthened me to get through the last, somewhat arduous, steps.

Heel veel dank ook aan enige vrienden: Maerten, Asia (en Louis), Jannes, Martin, Milan, Kristel, Joris, Rolf, Yannick, Sander, Tamas, Eka, Willem, Shamanee, Maris. Jullie toonden mij dat mijn zelfbeeld niet alleen hoeft te leunen op professionele prestaties, en zo vond ik de ontspanning om weer diep na te denken en hard te kunnen werken.

Papa, Peter, ik lijk steeds meer op jou. In veel opzichten heb ik mijn leven ingericht met het jouwe als blauwdruk. Jij en mam lieten me vrij om mijn eigen pad te kiezen, zolang ik mijn schouders er maar onder zette. Ik ben dankbaar dat ik zulke warme ouders mocht hebben. Kajan, Sylwia, Wilma, Marieke, onwijs veel dank voor de geweldige zorg voor Peter in deze laatste fase.

Dennis, Bob, Bert, Bram en Trudy — wat een rijkdom dat jullie zo dichtbij Lien (en Peter) zijn in het 'dorp' Groenburgwal. Door mijn basis te versterken, geven jullie mij ruimte om meer over T cellen en methodologie na te denken. Mama, zonder jouw steun had ik het waarschijnlijk niet tot de eindstreep gebracht. Wetenschap is tegenwoordig een teamsport, en met jouw inzicht aan mijn zijde ben ik een betere teamspeler.

CURRICULUM VITAE

Maarten Slagter was born on 27 October 1987 in Amsterdam. His fascination with biology and human physiology began in high school, alongside a deep passion for music. Aspiring to become a professional musician, he pursued jazz guitar at California State University, Northridge (2006–2007), supported by the Fulbright Campus Scholarship Program. Reflecting his diverse curiosity, he also undertook coursework in chemistry during this time. From 2008 to 2011, Maarten studied Biomedical Sciences at the University of Amsterdam, where he excelled at statistical methodology and bioinformatics. These interests led him to pursue a Master’s degree in Bioinformatics and Systems Biology at the University of Amsterdam and the Free University of Amsterdam (2011–2013), graduating *cum laude*. A pivotal moment came when he encountered a job advertisement by Joost Beltman for a PhD position in computational tumor immunology at the Netherlands Cancer Institute (NKI). He began his doctoral research in 2014 under Beltman’s supervision within the group of Ton Schumacher. When Beltman left to establish his own lab at the Department of Toxicology, Leiden Academic Centre for Drug Research (LACDR), Leiden University, Maarten recognized the unique opportunity to integrate wet and dry lab immunology at the NKI and strategically transitioned his PhD project. In 2015, he restarted his doctoral research under the joint supervision of Ton Schumacher and Lodewyk Wessels, continuing through to 2022. This thesis reflects the culmination of this work, which sits at the intersection of bioinformatics, systems biology, and T cell immunity in cancer. Building on this PhD foundation, Maarten now spearheads algorithm and software development towards the design of individualized cancer vaccines at the biotech company CureVac.

LIST OF PUBLICATIONS

References

1. Mezzadra, R. *et al.* Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature* **549**, 106–110 (2017).
2. Dijkstra, K. K. *et al.* Generation of tumor-reactive T cells by co-culture of peripheral blood lymphocytes and tumor organoids. *Cell* **174**, 1586–1598 (2018).
3. Slagter, M. *et al.* Benchmarking the Foreign Antigen Space of Human Malignancies. *Available at SSRN 3279415* (2018).
4. Scheper, W. *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nature medicine* **25**, 89–94 (2019).
5. Beck, R. J., Slagter, M. & Beltman, J. B. Contact-dependent killing by cytotoxic T lymphocytes is insufficient for EL4 tumor regression in vivo. *Cancer research* **79**, 3406–3416 (2019).
6. Voorwerk, L. *et al.* Immune induction strategies in metastatic triple-negative breast cancer to enhance the sensitivity to PD-1 blockade: the TONIC trial. *Nature medicine* **25**, 920–928 (2019).
7. Wellenstein, M. D. *et al.* Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. *Nature* **572**, 538–542 (2019).
8. Wells, D. K. *et al.* Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834 (2020).

-
9. McGrail, D. *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Annals of Oncology* **32**, 661–672 (2021).
 10. McGrail, D. *et al.* Validation of cancer-type-dependent benefit from immune checkpoint blockade in TMB-H tumors identified by the FoundationOne CDx assay. *Annals of Oncology* **33**, 1204–1206 (2022).
 11. Slagter, M. *et al.* Lack of detectable neoantigen depletion in treatment-naïve cancers. *bioRxiv*, 2023–06 (2023).
 12. Hoekstra, M. E. *et al.* Distinct spatiotemporal dynamics of CD8+ T cell-derived cytokines in the tumor microenvironment. *Cancer Cell* **42**, 157–167 (2024).