

# DECODING IMMUNE CHECKPOINT INHIBITOR RESPONSE

Biomarker Discovery and Predictive  
Modeling through Multi Omics Analyses

LILIAN MARIE BOLL

---

TESI DOCTORAL UPF / 2025

*Director de la tesi*

**Dr. Maria del Mar Albà Soler**

DEPARTAMENT DE MEDICINA I CIÈNCIES DE LA VIDA



Universitat  
Pompeu Fabra  
*Barcelona*



The project that gave rise to these results received the support of a fellowship from “la Caixa Foundation” (ID 100010434). The fellowship code is LCF/BQ/DI21/11860060.





## Acknowledgements

This thesis would not have been possible without the great support of the people around me, each of whom, in their own way, contributed to the completion of this work.

First and foremost, my deepest thank you to my supervisor, Mar, for her outstanding expertise and guidance, the countless meetings, and always-encouraging words. You are a true role model and the best mentor I could have wished for! A big thanks also goes to Jùlia. No matter how busy you were, you always made time for my questions and doubts.

I would like to express my gratitude to Robert Castelo for his insightful advice, and of course to Marta, Sergio and Joan Antoni - you turned every analysis into something genuinely enjoyable. A heartfelt thank you to the IT team, who supported me through endless hours of big data downloads, installation issues, and were there to spot the missing comma when I was about to surrender.

To everyone on the fourth floor -past and present- thank you so much for getting me through the doomy days. The laughter-filled lunch breaks, volley and ping-pong games, beers and parties. You made this journey truly unforgettable!

Finally, I would like to thank all those who supported me outside the lab. Your presence and encouragement means the world.

Ein riesiges Dankeschön an meine Familie, ohne deren Liebe und Unterstützung ich heute nicht der Mensch wäre, der ich bin.

Words cannot express how grateful I am for my incredible friends. Though we may live in different countries, my best friend Jana and my Glückszitronenquetsche Tribe, I could not imagine my life without

you. And Carli, no day is too dark for you to brighten it up with your sunshine. Thank you for always being there.

Last but by no means least, my deepest gratitude goes to Albert. These past four years have not only marked my academic journey, but also the beginning of our life together - and I could not be more grateful. You have been my biggest supporter and my most honest critic, your presence has shaped this journey more than words can say.

## **Abstract**

Immune checkpoint inhibitors have revolutionized cancer therapy, yet only a small fraction of patients respond to treatment. This thesis investigates the molecular mechanisms driving immunotherapy response focusing on bladder cancer. We generated and analyzed a new dataset of advanced bladder cancer patients treated with immune checkpoint inhibitors. This allowed us to validate existing biomarkers and provided new evidence linking mutation clonality to improved immunotherapy response. In continuation, we integrated omics data of six advanced bladder cancer cohorts to conduct an extensive biomarker exploration, identifying key factors and novel markers associated with response. Building on these findings, we developed a robust prediction model for immunotherapy response, achieving high accuracy and stability when validated in an independent cohort. Finally, we conducted an in-depth analysis of stop-loss mutations as one of the novel biomarkers discovered in the second part. We constructed a comprehensive database of stop-loss mutations from thousands of cancer patients to investigate their function and therapeutic potential. Collectively, this thesis advances our understanding of tumor-immune interactions and contributes to biomarker discovery in immunotherapy.



## Resum

Els inhibidors del punt de control immunitari han revolucionat la teràpia contra el càncer. Tot i això, només una petita fracció dels pacients respon al tractament. Aquesta tesi investiga els mecanismes moleculars que impulsen la resposta a la immunoteràpia, centrant-se en el càncer de bufeta. Hem generat i analitzat un nou dataset de pacients amb càncer de bufeta avançat que han estat tractats amb inhibidors del punt de control immunitari. Això ens ha permès validar els biomarcadors ja existents i aportar noves evidències que relacionen la clonalitat de les mutacions amb una millor resposta a la immunoteràpia. A continuació, hem integrat dades òmiques de sis cohorts avançades de càncer de bufeta per dur a terme una exploració exhaustiva de biomarcadors, identificant factors clau i nous marcadors associats amb la resposta al tractament. A partir d'aquests resultats, hem desenvolupat un model de predicció per a la resposta d'immunoteràpia, assolint una gran precisió i estabilitat en validar-lo amb una cohort independent. Finalment, hem dut a terme una anàlisi en profunditat de les mutacions stop-loss, un dels nous biomarcadors descoberts a la segona part de l'estudi. Per això, hem construït una base de dades completa de mutacions stop-loss de milers de pacients amb càncer, amb l'objectiu d'investigar-ne la funció i potencial terapèutic. En conjunt, aquesta tesi contribueix a aprofundir el coneixement de les interaccions entre el tumor i el sistema immunitari i impulsa el descobriment de nous biomarcadors en immunoteràpia.



## Zusammenfassung

Immuncheckpoint-Inhibitoren haben die Krebstherapie grundlegend verändert, jedoch spricht nur ein kleiner Teil der Patienten auf diese Art der Behandlung an. In der vorliegenden Doktorarbeit untersuchten wir molekulare Mechanismen, die das Ansprechen auf eine Immuntherapie beeinflussen, und legten dabei einen besonderen Fokus auf Blasenkrebs. Im ersten Schritt erstellten und analysierten wir einen neuen Datensatz von Patienten mit fortgeschrittenem Blasenkrebs, die mit Immuncheckpoint-Inhibitoren behandelt wurden. Dies ermöglichte uns, bereits bestehende Biomarker zu validieren und lieferte neue Beweise für den Zusammenhang zwischen Mutationsklonalität und einer verbesserten Antwort auf die Immuntherapie. Anschließend integrierten wir Omics-Daten von sechs Blasenkrebskohorten, um eine umfassende Biomarkeruntersuchung durchzuführen. Dabei wurden Schlüsselfaktoren und neue Marker identifiziert, die mit einem Ansprechen auf die Immuntherapie in Verbindung stehen. Abschließend führten wir eine eingehende Analyse von Stop-Loss-Mutationen mit im zweiten Teil identifizierten Biomarkern durch. Hierzu erstellten wir eine umfassende Datenbank mit Stop-Loss-Mutationen von Tausenden Krebspatienten, um deren Funktion und therapeutisches Potenzial zu erforschen. Insgesamt trägt diese Arbeit zu einem verbesserten Verständnis der Interaktionen zwischen Tumor und Immunsystem sowie zur Entdeckung von neuen Biomarkern für die Immuntherapie bei.





## Preface

Cancer remains one of the greatest challenges of modern medicine, not only as a biological puzzle asking to be solved but also as a profound human struggle. The discovery of immune checkpoint inhibitors has offered new hope to patients for whom traditional therapies have failed. Yet, the reality remains that only a fraction of individuals benefits, leaving the pressing question:

Why do some respond while others do not?

This thesis is driven by the conviction that science must bridge this gap: transforming uncertainty into understanding and data into actionable insights. Like many, I have known people diagnosed with cancer, and these experiences have reinforced my belief that research must serve the people behind the numbers.

Science, at its core, is an effort to bring light to the unknown. I hope this work, in its own small way, helps move us closer to a world where fewer conversations about cancer end in uncertainty.



# Contents

Acknowledgements	v
Abstract	vii
Resum	ix
Zusammenfassung	xi
Preface	xiii
1 INTRODUCTION	1
1.1 Cancer genome instability and the generation of neoantigens . . . . .	1
1.2 Genomic-based biomarkers of immunotherapy response . . . . .	4
1.3 Gene expression biomarkers of immunotherapy response . . . . .	8
1.4 Molecular subtypes in bladder cancer . . . . .	13
2 OBJECTIVES	19
3 RESULTS	23
3.1 The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer . . . . .	23
3.2 Predicting immunotherapy response in advanced bladder cancer through a meta-analysis of six independent cohorts . . . . .	39
3.3 Stop-loss mutations in cancer . . . . .	55
4 DISCUSSION	85
5 CONCLUSIONS	97
Bibliography	99
Supplementary Material	123
Annex	175



## 1.1 Cancer genome instability and the generation of neoantigens

Genome instability and immunity are the underlying drivers of human cancer (Hanahan and Weinberg, 2011). A healthy body cell accumulates somatic mutations increasing its proliferative capabilities, so called driver mutations, or passenger mutations that have no direct effect on tumor expansion. Somatic mutations can have a direct effect on the translated protein. The majority of non-synonymous mutations are missense mutations where a single nucleotide variant changes the encoded amino acid. Similarly, insertions and deletions will affect the protein sequence and can cause a frameshift. A stop-loss, or nonstop, mutation alters an initial stop codon to a coding amino acid which results in the continuation of the translation into the three prime untranslated region (3'UTR) of the mRNA. Copy number changes are another source of genetic alteration, where modifications in the DNA structure lead to the amplification or deletion of genetic sections. In contrast, synonymous mutations, or silent mutations, will not change the amino acid sequence.

A subset of somatic mutations is estimated to generate tumor-specific antigens. These antigens originate from the degradation of mutated protein sequences into peptides that can be loaded onto human leukocyte antigen (HLA) molecules, also known as major histocompatibility complex (MHC) in other animals. The MHC-peptide complex is then presented on the cancer cell surface. These presented neoantigens can be recognized by a T cell receptor (TCR) and induce T cell activation.

### 1.1.1 The hide-and-seek between tumor and immune cells

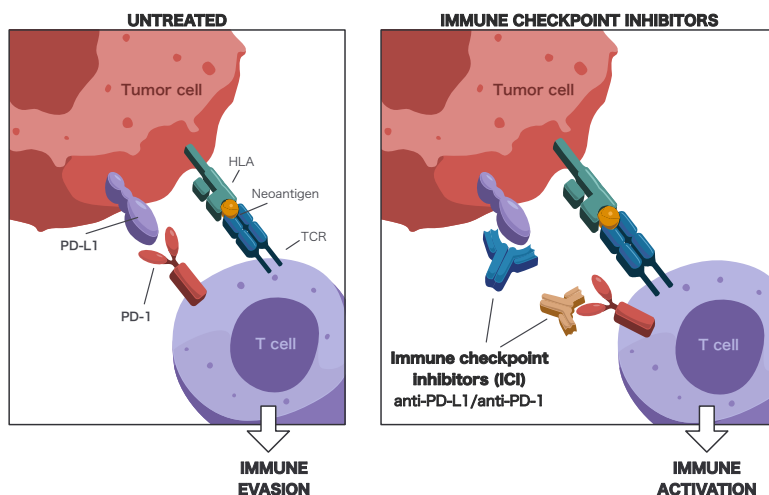
To avoid being detected and killed by the immune system, tumor cells upregulate the expression of immune checkpoint proteins on their surface. Immune checkpoints are an essential control measurement of the immune system, to prevent an overreaction of immune cells destroying healthy body cells. A ligand-receptor connection between a ligand on the body cell and its receptor on the T cell inhibits the activation of the latter. Tumor cells profit from this mechanism. By overexpressing the immune checkpoint proteins, they inhibit the T cell-induced immune response that would usually be triggered by a TCR recognizing a tumor neoantigen.

Immunotherapy describes treatment forms that enhance the patient's immune system to detect and kill cancer cells. The introduction of immunotherapy over a decade ago reshaped the treatment landscape drastically, with immune checkpoint inhibitors (ICIs) marking a significant breakthrough.<sup>1</sup> ICIs physically block the ligand-receptor interaction of immune checkpoints between the cancer and the T cell (Figure 1). The most prominent targeted immune checkpoints are the programmed cell death protein (PD-1) and its ligand, programmed cell death protein-ligand (PD-L1), or the cytotoxic T lymphocyte-associated antigen 4 (CTLA-4).

Treatment with ICIs leads to remarkable clinical response and long-term benefits in patients with different cancers, among others melanoma, urothelial carcinoma, renal cell carcinoma, head and neck squamous cell cancer, and non-small cell lung cancer (Wojtukiewicz et al., 2021). However, only a subset of patients show clear treatment response, and the treatment comes with possible immune-related adverse effects (Andres et al., 2022; Hussain et al., 2023; Sullivan and Weber, 2022; Yin et al., 2023), the most common being skin toxicities. The strong response in a small fraction of patients in combination with the potential toxicity of the treatment highlights the need for biomarkers

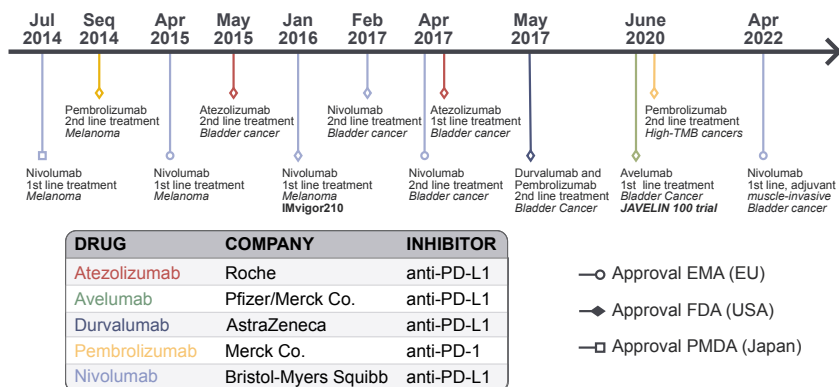
---

<sup>1</sup>Not least illustrated by the 2018 Nobel Prize in Medicine for immunologists James Allison and Tasuku Honjo.



**Figure 1: The effect mechanism of Immune checkpoint inhibitors (ICIs).** Tumor cells that overexpress PD-L1 can evade immune surveillance by binding to the PD-1 receptor on T cells, suppressing their activation and cytotoxic function (left). ICIs targeting PD-1 or PD-L1 disrupt this interaction, effectively reactivating T cells (right). HLA Human leukocyte antigen, PD-1 Programmed cell death protein, PD-1L Programmed cell death protein ligand, TCR T cell receptor.

separating responding patients from non-responders. Since the first authorizations of the anti-CTLA-4 drug ipilimumab in 2011 and the anti-PD-1 drug pembrolizumab in 2014 (Figure 2), scientists around the world sought to identify the variables driving ICI response to better predict treatment outcomes in the future.



**Figure 2: Timeline of approval for PD-L1 PD-1 inhibitors by the EU, USA and Japan.** Showing the first drug approved for first and second line treatment for all cancer types, bladder cancer, and high-TMB cancer. Adapted from the Cancer Research Institute, accessed December 17th, 2024 (Administration, 2025). EMA European Medicine Agency, FDA Food and Drug Administration, PD-1 Programmed cell death protein, PD-1L Programmed cell death protein ligand, PMDA Pharmaceuticals and Medical Devices Agency.

## 1.2 Genomic-based biomarkers of immunotherapy response

The rationale for the success of ICI is the reactivation of the immune system. T cells recognize neoantigens on the tumor surface. Tumor cells partially evade such T cell recognition by overexpression of immune checkpoints. In the presence of ICIs, the tumor cell cannot escape the consecutive immune response through activation of the immune checkpoint. As neoantigens result from mutations, a tumor accumulating many somatic mutations is more likely to be recognized by a T cell as non-self. The most relevant biomarker in the context of ICI response is, therefore, tumor mutational burden (TMB), defined as the number of mutations per DNA megabase (Mut/Mb). A high TMB is related to the generation of neoantigens, increasing the probability of them being presented by an MHC and being immunogenic. An association between high TMB and anti-CTLA-4 and later anti-PD-1/anti-PD-L1 response was first seen in melanoma, non-small-cell lung cancer, and bladder cancer (Hugo et al., 2016; Rizvi et al., 2015;



Rosenberg et al., 2016; Snyder et al., 2017; Van Allen et al., 2015). Samstein *et al.* showed TMB across different cancer types that a TMB in the upper quintile (20%) positively relates to overall survival across different cancer types (Samstein et al., 2019). In June 2020, the FDA approved 10Mb/Mut as a TMB-high threshold for pembrolizumab, based on a gene panel (for Drug Evaluation and Research, 2020).

While these cases show that overall patients with a high TMB tend to respond better, such a TMB threshold does not offer a one-fits-all solution to predict ICI response (Jung et al., 2023; McGrail et al., 2021; Usset et al., 2024). In many cases, the sole use of TMB is not sufficient to explain the response to ICI therapy (Samstein et al., 2019; Snyder et al., 2017). This was especially found for glioma, breast and prostate cancer, where neoantigen load does not positively correlate with CD8 T cell infiltration (McGrail et al., 2021), but also for high mutational tumors such as bladder cancer, where low TMB can show unexpected good clinical response while a patient with high TMB might not respond as expected.

Both selection pressure and genetic drift drive the accumulation of mutations in cancer. Clonal mutations occur at an early stage of cancer evolution and are found in essentially all tumor cells. In different studies, clonal TMB has been shown to add to the separation of ICI responders from non-responders compared to total TMB (Litchfield et al., 2021; McGranahan et al., 2016; Miao et al., 2018; Snyder et al., 2017), while in other cases the effect of total TMB was stronger (Usset et al., 2024). High tumor heterogeneity is expected to lead to a more dissimilar neoantigen load. Studies in mice have shown that with a high number of subclonal mutations, the immune response is less directed. Even in cases of highly immunogenic peptides, a substantial proportion of cells need to present them to trigger immune activation (Gejman et al., 2018; Wolf et al., 2019).

In the same way as TMB, neoantigen load relates to better response to ICIs (Richman et al., 2019; Shapiro and Bassani-Sternberg, 2023). For a mutation to elicit an immune response, it needs to be processed by

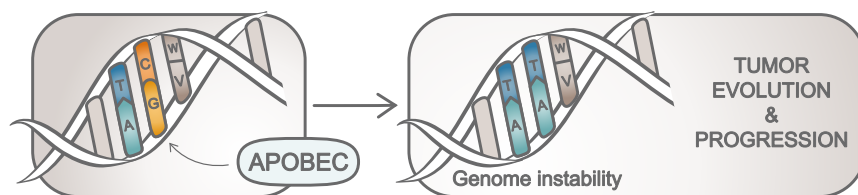
the tumor cell, loaded onto an HLA, and recognized by a TCR. Several algorithms have been developed to predict the binding affinity of a peptide to a patient's HLA type (Jurtz et al., 2017; O'Donnell et al., 2020). The number of these predicted neoantigens however strongly correlates with TMB and gives few new insights into response prediction (Mariathasan et al., 2018; Rech et al., 2018). In contrast to HLA binding, TCR recognition is difficult to predict. Instead, researchers try to predict the potential immunogenicity of a presented peptide. One hypothesis is that the higher a neoantigen's dissimilarity is to the host proteome, the more likely it is to be recognized by the immune system. This is the case for example when the mutated peptide is likely to be presented by the HLA, while the non-mutated counterpart is not (Rech et al., 2018; Richman et al., 2019).

In many cancer types, the majority of non-synonymous mutations are missense mutations, where mutated and wild-type sequences differ by a single amino acid. In contrast, insertions and deletions can shift the open reading frame generating completely new peptides previously unseen by the immune system. These distinct peptides are hypothesized to have a high potential for antigen recognition by the TCR due to the reduced likelihood of evicting self-tolerance mechanisms (Linnebacher et al., 2001; Turajlic et al., 2017). Indeed, frameshift neoantigens have been linked to ICI response in melanoma, lung, renal, and unstable colorectal cancer (Lindeboom et al., 2019; Maby et al., 2016). In the same way, the translation of non-coding regions poses a new source of potential tumor-specific neoantigens that can activate T cells, as shown in melanoma, lung or liver cancer (Camarena et al., 2024; Kraemer et al., 2023; Rogel et al., 2011). Similarly, a stop-loss mutation, where a terminating codon mutates to a coding one, can lead to continued translation into the 3'UTR. We hypothesize that such C-terminal peptide extensions constitute an additional source of immunogenic neoantigens expressed by tumor cells.

Frameshift mutations often cause premature stop codons and are therefore likely to be degraded at the messenger RNA level through nonsense-mediated decay (Litchfield et al., 2020). On the other hand, cases of stop-loss mutations without a downstream stop codon are digested by the nonstop decay surveillance pathway (Klauer and van Hoof, 2012).

### 1.2.1 Mutational signatures have been related to response.

The patterns of the trinucleotide context of a mutation can give information on the tumor evolution and its underlying driving forces. Many mutational signatures have been connected to biochemical processes, intrinsic (DNA damage repair, methylation, oxidation. . . ), and external (UV radiation, smoking. . . ) factors. In bladder cancer, the activation of apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of proteins has been associated with genomic instability and T>G mutations in a TWC motif (**Figure 3**) (Glaser et al., 2017; Weinstein et al., 2014). By the hypermutation of cytosines, *APOBEC3A* and *APOBEC3B* lead to an increased mutational burden and are believed to induce oncogenesis (Chan et al., 2015; Swanton et al., 2015).



**Figure 3: Schematic representation of APOBEC-induced mutations.** The DNA sequence TCW has been altered to TTW, a characteristic mutation pattern attributed to APOBEC cytidine deaminase activity. W-V is a placeholder for an arbitrary nucleotide pair.

## 1.3 Gene expression biomarkers of immunotherapy response

### 1.3.1 Immune infiltration

Cancer cells continuously alter their surrounding, the tumor microenvironment (TME). The TME consists of the extracellular matrix, blood vessels, immune cells, fibroblasts, and mesenchymal stromal cells that interact with and develop around a growing tumor. By activating fibroblasts, angiogenesis (the abnormal growth of vessels), and secretion of cytokines, cancer cells create an immunosuppressive surrounding. The composition and characteristics of the TME have a direct impact on prognosis as well as the patient's sensitivity to treatments. Especially in the case of immunotherapy, the immune system's ability to infiltrate the cancer tissue is a central condition since the recognition of a presented tumor antigen by the TCR requires physical proximity of the T cell to the tumor.

A stratification of tumors was proposed based on their level of immune-inflammation or immune-desertion into 'hot' or 'cold' (Gajewski, 2015). Later, an additional in-between state, the immune-excluded TME, was proposed (Hegde et al., 2016). The three immune phenotypes mainly differ by their state of T cell infiltration. An immune-infiltrated, 'hot' tumor has a high presence of active T cells inside the tumor tissue which has been repeatedly associated with an improved response to ICI in ovarian, triple-negative breast cancer, and non-small cell lung cancer (Desbois et al., 2020; Hammerl et al., 2021; Rosenthal et al., 2019). In the case of the immune-excluded phenotypes, T cells are only present in the peripheral tissues of the cancer, either due to missing stimuli or because they are physically blocked by stromal cells, which leads to a decreased response to ICI. Finally, a 'cold', or immune-deserted tumor lacks T cells both in the tumor tissue and its surroundings, limiting the initiation of the immune system in response to ICI, and relates to treatment resistance in melanoma, triple-negative breast, and ovarian cancer (Desbois et al., 2020; Hammerl et al., 2021).

Although associated with ICI response, this classification into immune phenotypes simplifies the complexity of the TME. To unravel how the TME affects immunotherapy, we must dive into its factors and their interactions in more detail.

Infiltration and inflammation of CD8+ T cells have been repeatedly associated with improved response to ICIs (Litchfield et al., 2021; Mariathasan et al., 2018). Different markers for CD8+ T cell infiltration have been presented, among them the expression of *CD8A* that strongly correlates with T cells in solid cancers (Dangaj et al., 2019). The migration of immune cells, especially of cytotoxic T cells, into and within the tumor tissue is orchestrated by pro-inflammatory chemokines, such as CXCL9/10/11, CCL5, and CXCR3 (Gorbachev et al., 2007). These molecules are at least partially initiated by another important proinflammatory cytokine, interferon-gamma (IFN- $\gamma$ ) that is secreted by lymphocytes in a pro-inflammatory state (Ayers et al., 2017; Cristescu et al., 2018). Through a positive feedback loop, IFN- $\gamma$  enhances its own production in T cells. By activating JAK-STAT1 (Janus kinase/signal transducers and activators of transcription) signaling, the chemokine creates an unfavorable TME that induces cancer cell apoptosis. Further, IFN- $\gamma$  drives the expression of PD-L1 in cancer cells, the maturation of naïve T cells to CD8+ T cells, and inhibits the immune-suppressive regulatory T cells (Ayers et al., 2017). Additionally, IFN- $\gamma$  stimulates the polarization of M0 macrophages to the anti-tumoral M1 phenotype and enhances their chemokine production (Jorgovanovic et al., 2020). Finally, by destructing endothelial cells and down-regulating vascular endothelial growth factor (VEGF), it negatively affects tumor angiogenesis.

In addition to the presence of CD8+ T cells in the TME, the expression level of the immune checkpoints is an important marker for ICI response. A patient is unlikely to respond to a treatment blocking the interaction of PD-L1 if the protein is barely expressed by the tumor cells (Balar et al., 2017; Chung et al., 2019; O'Donnell et al., 2019; Schmid et al., 2018). Consequently, the evaluation of PD-L1 status

obtained from immunohistochemistry assays has been approved by the FDA for urothelial cancer among other cancer types (Administration, 2025). Nevertheless, patients cannot be stratified perfectly by PD-L1 expression into responders and non-responders. The protein is expressed by different actors in the TME and expression values can be heterogeneous in the metastases, throughout cancer evolution, and affected by previous treatments (Mansfield et al., 2016; McLaughlin et al., 2016; Wang et al., 2019).

Besides CD8+ T cells, the TME contains different cell populations of the adaptive and innate immune system. Similar to the cytotoxic CD8+ T cells, supporting CD4+ T cells and B cells express PD-1 (Keir et al., 2008). In comparison to CD8+ T cells, the role of CD4+ T and B cells is less clear. Both secrete pro-inflammatory cytokines such as IFN- $\gamma$  that have a direct anti-tumor activity. A majority of CD4+ T cells are T helper (Th) cells that support CD8+ T cells by the secretion of pro-inflammatory chemokines, like interleukin 2 (Wang and Livingstone, 2003). In the TME, B cells are less likely to infiltrate the tumor but rather stay in close proximity, building tertiary lymphoid structures, presenting antigens to T cells, and secreting inflammatory cytokines such as IFN- $\gamma$  (Cabrita et al., 2020; Harris et al., 2000). At the same time, B cells can produce pro-tumor chemokines and their presence has been related to a worse prognosis. Over different cancer types, increased B cell content in the TME has been associated with ICI response (Cabrita et al., 2020; Helmink et al., 2020; Kim et al., 2021; Petitprez et al., 2020; Schaafsma et al., 2021). Then again, other research groups report no association with treatment response (Damsky et al., 2019).

B and T cells are part of the adaptive immune response, but also cells of the innate immune system, the body's nonspecific defense mechanism, modulate the TME. An important player is tumor-associated macrophages. Depending on the stimuli, monocyte-derived macrophages can differentiate into pro-inflammatory M1 macrophages or wound-healing, anti-inflammatory M2 macrophages that show im-

munosuppressive activities, promoting angiogenesis and stromal cell and cancer cell growth (Cassetta and Pollard, 2018). M1 macrophages primarily secrete chemokines to attract cytotoxic lymphocytes and create an immune-stimulating environment, but can also kill cells by intrinsic phagocytosis (Dangaj et al., 2019; House et al., 2020).

### 1.3.2 Immune suppression

While the immune system is often thought to create an anti-tumor setting, some players of the immune system can strengthen the pro-oncogenic TME. Regulatory T (Tregs) cells, a type of CD4<sup>+</sup> T cells, modulate the host's immune response to prevent autoimmunity and chronic inflammation. By suppressing T cell proliferation and cytokine secretion in inflammatory sites, they form an important part of the immune escape mechanism (Togashi et al., 2019). Their secretion of growth factors directly supports the growth of cancer cells as well as fibroblasts and endothelial cells. Such anti-inflammatory cytokines are the transforming growth factor beta (TGF- $\beta$ ) and interleukin 10 (IL-10) (Togashi et al., 2019). Especially TGF- $\beta$  has been intensively studied in relation to immunotherapy response due to its ability to exclude CD8<sup>+</sup> T cells from the TME (Mariathasan et al., 2018). Consequently, patients with low T cell infiltration profiles due to high TGF- $\beta$  show low response rates to ICI. Moreover, TGF- $\beta$  is expressed by the anti-inflammatory M2 macrophages, stimulating the conversion of CD8<sup>+</sup> T effector cells to anti-inflammatory Tregs.

Finally, TGF- $\beta$  is a crucial player in the generation of cancer-associated fibroblasts by epithelial-mesenchymal transition (EMT). Stroma cells, such as endothelial cells, fibroblasts, or lipocytes, build a central part of the TME. Once the tumor surpasses a size where it can rely on passive diffusion and before entering hypoxic and acidic conditions, it will build out its blood vessels for nutrition, oxygen supply, and waste management. The transition of endothelial cells to cancer-associated fibroblasts further allows the tumor to migrate and initiate metastases (Jorgovanovic et al., 2020). In this context, EMT-related

gene expression has been suggested as a source of immune resistance in urothelial cancer and is associated with worse response to ICI therapy (Wang et al., 2018).



## 1.4 Molecular subtypes in bladder cancer

Cancer tissue can be highly heterogeneous, which makes it difficult to come up with a one-fits-all therapy approach. Molecular subtypes are an attempt to group cancer patients into subgroups based on genetic mutations, gene expression, or other biomarkers. Especially in breast cancer, molecular subtypes have shown great clinical significance in creating treatment standards (Goldhirsch et al., 2013). As a high-mutational cancer, ICI was approved early for bladder cancer patients. Nevertheless, it remains the costliest cancer to treat (Michaeli et al., 2022). Bladder cancer separates into non-muscle invasive and muscle-invasive bladder cancer, with the latter covering less than 30% of the cases but being more aggressive, especially when metastatic (5-year survival rate <10%). After molecular classification of breast cancer tumors was successfully related to treatment outcome and led to clinical implications, many attempts were made to propose similar subtyping for bladder cancer samples based on their transcriptomic and mutational profiles (Höglund et al., 2023; Kamoun et al., 2020; Robertson et al., 2017). Usually, clusters for subtypes are obtained from hierarchical clustering, where each study developed its classifier on a different patient population. While there is no agreement on the exact clusters, different groups repeatedly divide the samples into subgroups within the two supersets basal and luminal. One of the classifiers, developed on the cancer genome atlas (TCGA) bladder cancer cohort defined five clusters:

1. Basal-squamous
2. Luminal-infiltrated
3. Luminal-papillary
4. Luminal
5. Neuronal

The basal-squamous subtype shows a strong expression of basal and stem markers such as *CD44* and *KRT5/6A/14*, as well as a strong immune inflammation signature, it has been repeatedly related to a positive response to immunotherapy (Kamoun et al., 2020; Robertson et al., 2017). Similarly, the luminal-infiltrated subtype is defined by a high infiltration of lymphocytes as well as a strong expression of uroplakins and EMT markers, and it was found to relate to response in one study (Robertson et al., 2017). In another urothelial cohort, only the two least common subtypes luminal and neuronal subtype were related to response to the anti-PD-L1 atezolizumab (Kim et al., 2019). Neuronal is characterized by neuroendocrine differentiation and shows poor survival rates, consistent with the high aggressiveness of neuroendocrine bladder cancer (Robertson et al., 2017).

#### 1.4.1 Prediction models for ICI response

The development of predictive models for therapy outcomes represents a critical step toward the realization of personalized medicine in oncology. As discussed in the previous paragraphs, tumor is a heterogeneous disease and a patient's response to immunotherapy has high variability. In addition, immunotherapies are often expensive and can have severe adverse effects. Reliable models to predict treatment responses could give valuable insights into underlying biological processes and eventually enable clinicians to tailor therapies to individual patients to improve overall outcomes. The field of medical prediction models profited from recent progress in two fields. Firstly, the advancements in next-generation sequencing techniques allow researchers to generate large-scale data faster at a decreasing financial cost. In this regard, research can take advantage of data sharing initiatives such as TCGA, or other international consortia and collaborations, making the data available to other scientists. Secondly, the rapid development of machine learning and artificial intelligence has provided powerful tools to analyze complex, multi-omics data, increasing our capacity to build predictors beyond simple biomarker-based frameworks.

Pan-cancer studies combine data from multiple cancer types to leverage the power of large datasets. This approach has clear advantages, but it also comes with notable challenges. By aggregating data across cancer types, pan-cancer studies benefit from larger sample sizes, which are crucial for training and testing robust machine learning models. These studies can uncover shared molecular mechanisms or biomarkers across cancer types, which may point to fundamental principles of tumor biology and more general predictors of immunotherapy response. However, different cancer types often have distinct molecular and clinical characteristics. Tumor-specific factors are obscured by pooling data, leading to oversimplified models that fail to capture the heterogeneity of individual cancer types. Further, combining data from multiple cancer types can introduce confounding variables, such as varying clinical protocols, molecular features, or treatment regimens across cancers, which may bias the results.

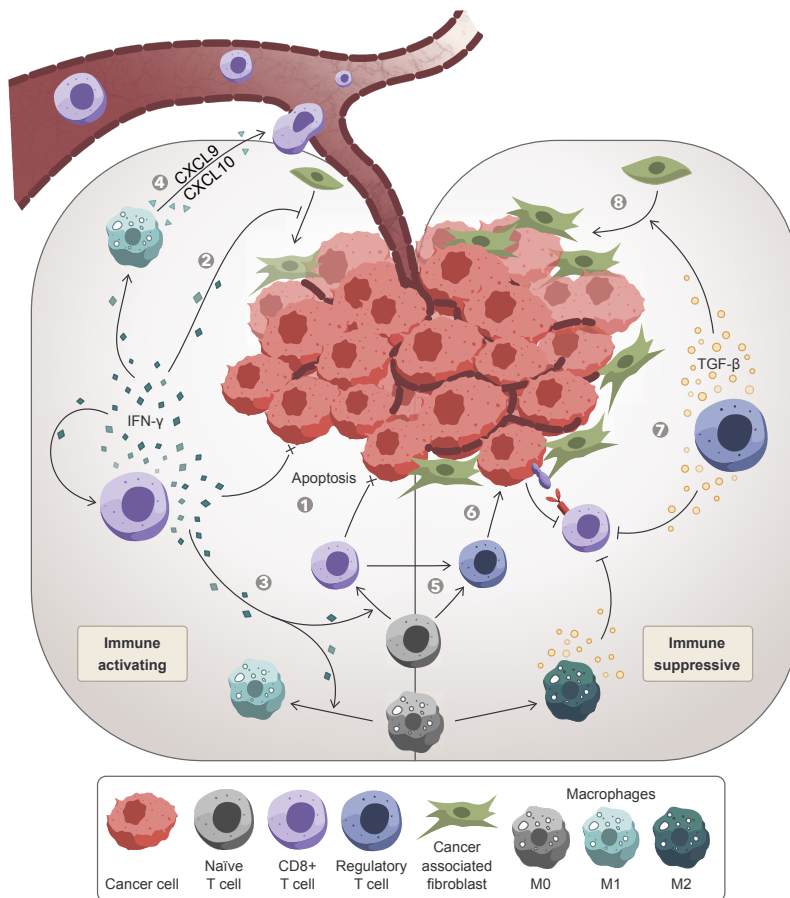
One big study from 2021 analyzed seven cancer types from over 1000 patients treated with different ICIs. The very large sample size and multi-omics data allowed them to outline some common features of the tumor antigen recognition and T cell initiation shared among cancer types (Litchfield et al., 2021). Depending on the selected features and test cohort, the AUC varies between 0.63 and 0.79 for a simple TMB and *CXCL9* model, and 0.66 to 0.86 for the full model. Another study of similar sample size focussed on mutation data to build a gene signature associated with overall survival to ICI (Long et al., 2022). Other groups related gene signatures obtained from expression data of smaller pan-cancer datasets to immunotherapy response (Benguigui et al., 2024; Hernando-Calvo et al., 2023; Long et al., 2022).

Focusing on a specific cancer type allows for more tailored analyses but comes with its limitations. These studies can uncover nuances and mechanisms specific to a given cancer type, providing more actionable insights for disease-specific treatment and management. This allows features included in the model to be more relevant and less diluted by irrelevant features from other tumor types, potentially improving






their predictive accuracy. With an increased number of studies on bladder cancer patients who received ICI treatment, groups built the first cancer-specific prognostic and prediction models using only this tumor type (Abuhelwa et al., 2021; Damrauer et al., 2022; Pond et al., 2018). The main drawback in cancer-specific studies is the amount of available data. Machine learning models require large datasets to be trained, tested and validated appropriately. Models limited to one cancer type are often developed on one or two datasets and frequently lack external datasets for validation, making it difficult to assess their generalizability and robustness. Small datasets therefore increase the risk of overfitting, where the model performs well on the training data but poorly on unseen data. While bladder cancer-specific prediction models have provided valuable biological insights, their ability to predict immunotherapy response depends strongly on the closeness of the test to the training cohort, reflected in the big range of AUC values (Damrauer et al., 2022).

Both pan-cancer and cancer-specific studies are influenced by the quality of the datasets. Differences in patient populations, sequencing platforms, and data preprocessing pipelines can lead to biases or irreproducible results. Batch effects can make it difficult to distinguish biological signals from technical artifacts. Furthermore, machine learning models are often complex and lack interpretability, making it difficult for clinicians to trust and adopt them in practice.

In an ideal world, models could be fed larger, more diverse datasets capable of capturing the subtle heterogeneity within patient populations, obtained from standardized data processing frameworks to minimize technical biases. While the clinical application of predictive models for immunotherapy remains distant, each iterative advance builds toward the ultimate goal of personalized oncology. By understanding the strengths and shortcomings of current approaches and existing tools, this thesis seeks to identify critical knowledge gaps and contribute to the foundation for further advancements to a more personalized cancer medicine.



**Figure 4: The role of immune cells in the tumor microenvironment depends on context and stimulation.** Besides tumor cells, solid cancers consist of a mixture of stromal cells and cells of the adaptive and innate immune system, inducing either an anti-tumor environment (left) or promoting tumor growth (right). Cytotoxic CD8+ T cells kill cancer cells upon recognition of tumor antigens (1). By secreting IFN- $\gamma$  they suppress angiogenesis and promote the differentiation of CD8+ T cells and M1 macrophages (2-3). M1 macrophages secrete pro-inflammatory cytokines like CXCL9/10 that attract more T cells to the tumor site (4). In the presence of cytokines such as TGF- $\beta$ , naïve and CD8+ T cells differentiate to regulatory T cells (5). These cells support tumor growth by secreting growth factors (6) and anti-inflammatory cytokines such as TGF- $\beta$  (7), which is also secreted by M2 macrophages. TGF- $\beta$  is further inducing the transition of endothelial cells to cancer-associated fibroblasts (8), enhancing tumor migration and angiogenesis. This figure simplifies the interactions of the tumor microenvironment, focusing on key components and omitting certain chemokines and immune cells.

	SUBTYPE	GENE EXPRESSION	COMMON MUTATIONS	OVERALL SURVIVAL	TREATMENT RESPONSE	IMMUNE INFILTRATION
Basal	<b>Basal-squamous</b>	Basal and stem-like Squamous differentiation Carcinoma in situ (CIS) signature Loss of SHH signalling Immune markers (including <i>PD-1</i> and <i>PD-L1</i> )	<i>TP53</i> mutation		Positive association to ICI (Kamoun et al., 2020; Sharma et al., 2017; Robertson et al., 2017) Positive association to cisplatin-based NAC (Sharma et al., 2017)	
	<b>Luminal-infiltrated</b>	Smooth muscle signatures Myofibroblast signatures Immune markers (including <i>PD-1</i> and <i>PD-L1</i> )		Negative association with OS (Kim et al., 2019)	Positive to ICI (Robertson et al., 2017) Resistant to cisplatin-based NAC (Seiler et al., 2017)	
Luminal	<b>Luminal-papillary</b>	Low carcinoma in situ (CIS) signature	<i>FGFR3</i> mutation and amplification <i>CDKN24</i> deletion	Best association with OS (Kim et al., 2019)	Negative to cisplatin-based chemotherapy (Seiler et al., 2017)	
	<b>Luminal</b>	Uroplakins Terminally differentiated urothelial umbrella cells		Negative association with OS (Kim et al., 2019)		
Neuronal	<b>Neuronal</b>	Neuroendocrine and neural development, Upregulation of p53 and cell-cycle	<i>TP53</i> and <i>RB1</i> mutations <i>E2F3</i> amplification	Poorest association with OS (Kim et al., 2019, )	Positive association to immune checkpoint inhibitors (Kim et al.)	

**Figure 5: The five TCGA subtypes in bladder cancer. Adapted from Robertson et al. (2017) and Kamoun et al. (2020).** The immune infiltration column indicates the state of immune infiltration in the subtypes, based on Robertson *et al.*.

The overall aim of my thesis is to deepen the biological understanding of the markers and drivers that lead to ICI response across various solid cancer types, with a particular focus on bladder cancer. In order to fulfill this aim, the following specific objectives are set:

**1. Integration of Omics Data from Tumors**

To set up a comprehensive pipeline for analyzing genomic and transcriptomic data from cancer patients.

**2. Biomarker Characterization**

To characterize and validate known biomarkers as well as unravel novel factors related to immunotherapy response, identifying the most relevant variables in the complex interplay between tumor biology, the immune microenvironment, and treatment response.

**3. Impact of Molecular Subtypes on ICI response**

To investigate the impact of bladder cancer molecular subtypes on the response to ICI treatment and study the interplay of reported biomarkers within these subtypes.

**4. Prediction Model Development**

To develop a robust bladder-cancer specific prediction model of immunotherapy response, integrating diverse sources of omics data.

**5. Stop-Loss Mutation Analysis**

To conduct an in-depth analysis of stop-loss mutations, their role in cancer development, and their potential to generate tumor-specific neoantigens.

Section 3.1 of this thesis combines the first and to some extent the second objective. Here I conduct a comprehensive analysis of a single advanced urothelial bladder cancer cohort treated with anti-PD-1/anti-PD-L1. While this single cohort allowed me to develop a comprehensive framework to analyze genomic-based and gene expression markers and provided valuable biological insights into ICI response in bladder cancer, it also shows that to achieve objective two, more data is needed.

Objective two is therefore further addressed in the second result section, together with objectives three and four. Building on the results of part one, this work combines and harmonizes six datasets to construct the currently largest bladder cancer-specific cohort. The comprehensive sample size enhances the statistical power, enabling the validation of several trends observed in the first part as statistically significant. It further allows the discovery of additional biomarkers such as the effect of stop-loss mutations and tumor-specific long-non-coding RNAs on ICI response, providing deeper insights into ICI response mechanisms. Furthermore, this part delves into the role of molecular subtypes in advanced bladder cancer, exploring their impact on treatment response and biomarker interplay. Finally, bladder cancer-specific prediction models for ICI response are developed and presented, integrating diverse sources of omics data to enhance predictive accuracy and clinical applicability.

The last part of my thesis addresses the potential role of stop-loss mutations in various cancer types. To approach this, I generated and downloaded available mutation data to extract stop-loss mutations. Profiting from the computational pipeline set up at the beginning of the project, I have predicted the C-terminal extensions resulting from stop-loss mutations and their potential to be presented by the tumor cell. Immunopeptidomics data served as experimental proof of the expression and processing of such stop-loss mutations.

By achieving these objectives, this thesis aims to provide novel insights into the mechanisms underlying ICI response, with a particular focus



on bladder cancer. The findings will contribute to a more nuanced understanding of tumor-immune interactions, inform biomarker discovery and validation, and pave the way for the development of clinically relevant predictive tools that can guide personalized immunotherapy strategies in cancer treatment.



### 3.1 The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer

**Full citation:**

Boll LM, Perera-Bel J, Rodriguez-Vida A, Arpí O, Rovira A, Juanpere N, Vázquez Montes de Oca S, Hernández-Llodrà S, Lloreta J, Albà MM, Bellmunt J. The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer. Sci Rep. 2023 Sep 15;13(1):15287. doi: 10.1038/s41598-023-42495-2. PMID: 37714872; PMCID: PMC10504302.

**First authors:** Lilian Marie Boll and Júlia Perera-Bel

**Status:** Published on September 15<sup>th</sup>, 2023

**Journal:** Scientific Reports

**Correspondence:** malba@researchmar.net and joaquim\_bellmunt@dfci.harvard.edu

**Supplementary Material:** Appendix, section 5.1.



# OPEN The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer

Lilian Marie Boll<sup>1,7</sup>, Júlia Perera-Bel<sup>1,7</sup>, Alejo Rodríguez-Vida<sup>1,2,3</sup>, Oriol Arpi<sup>1</sup>, Ana Rovira<sup>1,2,3</sup>, Núria Juanpere<sup>1</sup>, Sergio Vázquez Montes de Oca<sup>1</sup>, Sílvia Hernández-Llodrà<sup>4</sup>, Josep Lloreta<sup>1,4</sup>, M. Mar Albà<sup>1,5</sup>✉ & Joaquim Bellmunt<sup>1,6</sup>✉

Immune checkpoint inhibitors (ICI) have revolutionized cancer treatment and can result in complete remissions even at advanced stages of the disease. However, only a small fraction of patients respond to the treatment. To better understand which factors drive clinical benefit, we have generated whole exome and RNA sequencing data from 27 advanced urothelial carcinoma patients treated with anti-PD-(L)1 monoclonal antibodies. We assessed the influence on the response of non-synonymous mutations (tumor mutational burden or TMB), clonal and subclonal mutations, neoantigen load and various gene expression markers. We found that although TMB is significantly associated with response, this effect can be mostly explained by clonal mutations, present in all cancer cells. This trend was validated in an additional cohort. Additionally, we found that responders with few clonal mutations had abnormally high levels of T and B cell immune markers, suggesting that a high immune cell infiltration signature could be a better predictive biomarker for this subset of patients. Our results support the idea that highly clonal cancers are more likely to respond to ICI and suggest that non-additive effects of different signatures should be considered for predictive models.

## Abbreviations

AUC	Area under the curve
APOBEC	Apolipoprotein B mRNA editing catalytic polypeptide-like
BLCA	Bladder cancer
CCF	Cancer cell fraction
CNAG	Centro Nacional de Análisis Genómico, Barcelona
DAI	Differential agretopicity index
DE	Differential expression
DDR	DNA damage repair
FFPE	Formalin-fixed paraffin-embedded
GSEA	Gene set enrichment analysis
GTF	Gene transfer format
HLA	Human leukocyte antigen
ICI	Immune checkpoint inhibitor
MHC	Major histocompatibility complex
MSigDB	Molecular Signatures Database
NK	Natural killer
PD-1	Programmed death-1
PD-L1	Programmed death ligand-1
RECIST	Response Evaluation Criteria In Solid Tumors

<sup>1</sup>Hospital del Mar Research Institute, Barcelona, Spain. <sup>2</sup>Medical Oncology Department, Hospital del Mar, Barcelona, Spain. <sup>3</sup>Centro de Investigación Biomédica en Red de Oncología (CIBERONC-ISCIII), Barcelona, Spain. <sup>4</sup>Department of Medicine and Life Science, Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>5</sup>Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain. <sup>6</sup>Dana Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. <sup>7</sup>These authors contributed equally: Lilian Marie Boll and Júlia Perera-Bel. ✉email: MALba@imim.es; joaquim.bellmunt@gmail.com

RNA	Sequencing (RNASeq)
SB	Strong binder
TCGA	The cancer genome atlas
TGF- $\beta$	Transcription growth factor beta
TMB	Tumor mutational burden
VAF	Variant allele frequency
WES	Whole exome sequencing

Cancer is among the greatest burdens of disease in high- and middle-income countries, with urothelial cancer being the sixth most common cancer type in men globally<sup>1</sup>. Genetic aberrations play an important role in the development of urothelial cancer. Somatic variants resulting in mutated proteins can be processed and presented on the tumor cell's surface as small peptides called neoantigens. Neoantigens are tumor-specific, and thus, have a selective potential for T cell recognition<sup>2</sup>. One way tumors cells evict immune escape is by overexpressing checkpoint inhibitors. When binding to the ligand on the T cell surface, an immune response is suppressed by downregulating the production of cytokines, effector functions and T cell proliferation<sup>3</sup>. In urothelial carcinoma, high expression of the checkpoint molecule programmed death-ligand1 (PD-L1) has been associated with advanced cancer stages and poor survival<sup>4</sup>.

Immune checkpoint inhibitors (ICIs) have been a major advance in immunotherapy in the past decade. Monoclonal antibodies block immune checkpoints to prevent tumor cells from escaping T cell recognition. To date, several PD-1/PD-L1 inhibitors have been approved by the FDA and EMA, for patients with advanced urothelial cancer. However, less than one-third of urothelial cancer patients respond to ICI treatment<sup>5–8</sup>. Accurate methods to predict which patients are going to respond to the treatment are currently missing and are urgently needed in the clinics.

Based on the assumption that a higher tumor mutational burden (TMB) translates to an increased neoantigen load, TMB is nowadays one of the most relevant biomarkers currently being studied for clinical use in urothelial cancer<sup>9,10</sup>. However, findings are inconsistent across studies, and the isolated use of TMB has not been shown to accurately differentiate between responders and non-responders<sup>8,11,12</sup>. One possible way to improve the accuracy of the predictions is to consider neoantigen quality in addition to quantity<sup>13</sup>. Several computational prediction programs exist to estimate characteristics of neoantigens such as the ability to form a stable complex with the major histocompatibility complex (MHC) receptor or the probability of immune cell recognition. Yet, experimental studies showed that many of these computationally predicted epitopes are not presented on the cell surface<sup>14</sup>, limiting the applicability of these approaches.

Previous efforts to investigate the impact of different biomarkers in the response to ICI have mostly centered on melanoma and lung cancer—the cancers with the largest number of mutations—or in the meta-analysis of different types of cancer. These studies have suggested that the clonality of the mutations might be especially relevant. For example, McGranahan et al. showed that sensitivity to PD-1 and CTLA-4 blockade in patients with advanced non-small cell lung cancer and melanoma was enhanced in tumors enriched for clonal neoantigens<sup>15</sup>. These findings are consistent with Wolf et al. who report that clonal neoantigens relate to immune infiltration and clinical outcome in melanoma<sup>16</sup>. Another study including patients with urothelial cancer treated with ICIs targeting PD-1/PD-L1 and CTLA-4 found that clonal mutations were significantly higher in complete responders compared to non-responders or partial responders<sup>17</sup>. Finally, Litchfield et al.<sup>18</sup>, who performed a meta-analysis of biomarkers from different cancer types, found that clonal TMB was the strongest predictor of CPI response followed by total TMB and CXCL9 expression. Subclonal TMB, somatic copy alteration burden, and histocompatibility leukocyte antigen (HLA) evolutionary divergence failed to attain pan-cancer significance.

In order to gain knowledge on what might drive the response to monoclonal antibodies blocking the programmed cell death-1 (PD-1) or its ligand (PD-L1) in advanced urothelial cancer, and how this compares to other kinds of cancers, we performed an in-depth study of whole exome and RNA sequencing data from a cohort composed of 27 patients treated with ICIs targeting PD-1/PD-L1 from Hospital del Mar in Barcelona (Spain). As an independent dataset, we also analyzed somatic variation in urothelial cancers from a 25 patients cohort treated with the PD-L1 inhibitor atezolizumab at Memorial Sloan Kettering Cancer Center<sup>8</sup>. The results provide additional novel insights into the genomic correlates associated with immunotherapy response and suggest ways in which predictive models could be improved.

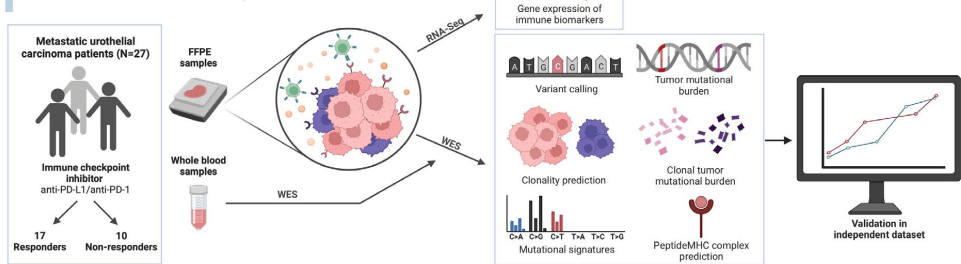
## Results

**Genomic analysis and cohort description.** We generated whole-exome sequencing (WES) and RNA sequencing (RNA-Seq) data from tumors and blood samples from 27 advanced urothelial cancer patients treated with anti-PD-1/PD-L1 ICIs at Hospital del Mar (Fig. 1a). WES data was obtained from the tumors before treatment as well as from blood samples, allowing reliable identification of cancer-specific mutations. The type of response to treatment was defined using the Response Evaluation Criteria In Solid Tumors (RECIST) criteria 1.1. Among the 27 patients whose tumors were sequenced, 17 were responders to the ICI treatment (12 partial, 5 complete), and 10 were non-responders. Complete response is described as the complete disappearance of the tumor tissue, partial response a decrease of the target lesion by at least 30%. Additionally, we obtained high-quality RNA-Seq data from tumor samples in 20 patients of which 13 were responders to the ICI treatment (8 partial, 5 complete responses) and 7 were non-responders (progressive disease). The mutational data was used to investigate the effect of the number and type of mutations in response to treatment, as well as to predict putative neoantigens on a patient basis. The gene expression data was used to investigate the impact of different immune-based markers in ICI response.

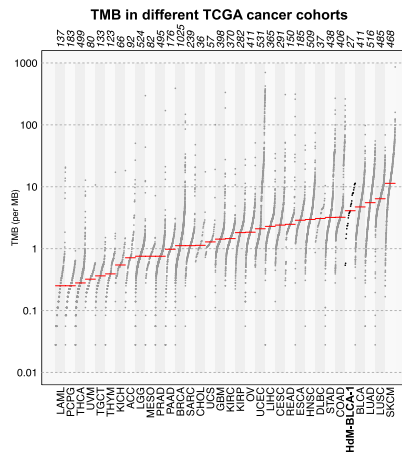
a

## STUDY DESIGN and OMICS DATA ANALYSIS

### Biomarkers of immunotherapy response in metastatic urothelial carcinoma patients

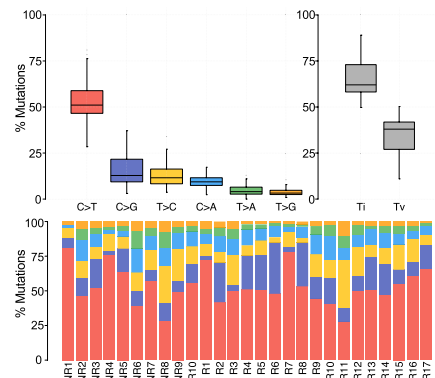


**b**



**C**

### Transition and Transversion in HdM-BLCA-1 cohort



d

		Misense	Nonsense	Frameshift	Number of Patients		Misense	Nonsense	Frameshift	Number of Patients		Misense	Nonsense	Frameshift	Number of Patients		Misense	Nonsense	Frameshift	Number of Patients			
Tumor suppressor genes						Tumor suppressor genes						Oncogenes						Oncogenes					
ELF3	3	1	4	8		ARID1B	1	1	2		PKRFD	6		6		KRAS	2			2			
KMT2D	7	1		8		ARID2		1	1	2		GFGR3	5		5		PIK3R1	2			2		
ARID1A	2	3	2	7		ASXL1	2			2		KDM6A	3	1	1	4							
TP53	6			6		ATM	2			2		EP300	2	2		4							
KMT2A	3	1	1	5		CDKN2A	2			2		ERBB3	4			4							
KMT2C	2	1	2	5		FBXW7	1	1		2		IRS4	3	1		4							
APC	4			4		FOXQ1	1		1	2		RHOA	4			4							
BRCAC2	3	1		4		KLF5	2			2		ERBB2	3			3							
RB1		1	3	4		PPARG	2			2		CHOB	3			3							
ASXL2	3			3		PtCH1	2			2		CND3	2			2							
ATRX	2	1		3		STAG2	2			2		CTNNB1	2			2							
BRCAC1	2		1	3		ZFP36L2	2			2		HRAS	2			2							
Tumor suppressor genes						Oncogenes						Oncogenes						Function unclear					
ZFP36L1						ZFP36L1						ZFP36L1						FAM47C					
FAM47C						FAM47C						FAM47C						ZFP36L1					

**Figure 1.** Analysis of omics data from the metastatic urothelial cancer cohort at Hospital del Mar. (a) Study design. We performed whole exome sequencing (WES) and RNA sequencing (RNA-Seq) from a cohort of 27 patients. The cohort included 17 responders (partial or complete response) and 10 non-responders (progressive disease). We measured the differences in the distribution of several biomarkers, including clonal and subclonal TMB, mutational signatures, number of predicted mutation-induced neoantigens and gene expression signatures in responders and non-responders. Created with BioRender.com. (b) Tumor mutational burden (TMB) compared to different TCGA cancer cohorts. The number of non-synonymous mutations in the cohort in our cohort (HdM-BLCA-1) was very similar to the bladder cancer TCGA cohort and in accordance with this being a highly mutated cancer. The names below the plot represent the TCGA cohort abbreviation, the numbers above represent the cohort sizes. (c) Frequencies of different pairwise nucleotide substitutions. C->T mutations were the most frequent ones, followed by C->G. TT transition, Tv transversion. (d) Cancer-related frequently mutated genes. Missense, frameshift and nonsense mutations are shown. Genes have been classified into 'Oncogenes', 'Tumor suppressor genes' and 'Function unclear' based on previous knowledge. NR no responders, R responders.

**Estimation of tumor mutational burden and identification of recurrently mutated genes.** Overall, we identified 6989 cancer-specific non-synonymous mutations. Most mutations were missense, 6192 (88.6%), but we also detected 490 nonsense (premature stop codons) and 130 frameshift mutations. We calculated the TMB for each patient as the sum of non-synonymous mutations. The average TMB of the dataset was 5.18 mutations per Mb, comparable to that observed in the cancer genome atlas (TCGA) bladder cancer cohort (Fig. 1b). Among pairwise nucleotide substitutions, C->T mutation was the most common (Fig. 1c), with a median frequency of around 0.5, very similar to that observed in the TCGA bladder cancer dataset (Figure S1) (18). The second most common mutation was C->G. We also specifically searched for the APOBEC signature, which has been previously found to be frequent in urothelial cancer<sup>19,20</sup>. APOBEC cytidine deaminases have the ability to introduce mutations in chromosomal and mitochondrial DNA potentially driving oncogenesis<sup>21</sup>. In the APOBEC signature C mutates to G or T in the context of TCW motifs (where W is A/G). Using the maftools package<sup>22</sup>, we found that this signature was significant in 16 out of 27 sequenced tumors.

We identified previously described driver and tumor suppression genes that were mutated in more than one patient in our cohort. These included genes in pathways related to TP53/cell cycle (*TP53*, *RBI*, *ATM*), histone modification (*KDM6A*, *EP300*, *KMT2A/C/D*), DNA damage (*BRCA1/2*) and chromatin remodeling (*ELF3*, *ARID1A*, *ARID1B*, *ARID2*)<sup>19,23,24</sup> (Fig. 1d). The two most mutated genes were *ELF3* and *KMT2D*, with alterations in 8 patients. *ELF3* is an important transcriptional regulator for the differentiation of the urothelium with a high mutation frequency in bladder cancer<sup>25</sup>, and histone-lysine *N*-methyltransferase genes (*KMT2A/C/D*) are among the most mutated in different cancer types<sup>26,27</sup>. The majority of changes in *ELF3* were frameshift and nonsense mutations, similar to the findings of Nordentoft et al.<sup>24</sup>. In contrast, in *KMT2D*, 7 out of 8 mutations were missense.

**Tumor mutational burden is significantly associated with the response to ICI.** Next, we investigated the relationship between different types of mutations and the response to treatment. We found that the number of non-synonymous mutations or TMB was significantly higher in responders than in non-responders (Fig. 2a, Wilcoxon test  $p$  value = 0.046). This is in agreement with previous observations for different cancer types<sup>17,18,28–30</sup>. Additionally, we found that the number of nonsense mutations was also higher in responders than in non-responders (Fig. 2b). Responders also tended to have more frameshift mutations than non-responders, although in this case, the difference was not statistically significant (Fig. 2c).

**Clonal TMB better separates responders from non-responders than total TMB.** Some mutations in the tumor are present in all cells (clonal), whereas others are only present in a subset of the cells (subclonal). We used the cancer cell fraction (CCF) statistic, in which the initial allele frequencies are corrected by tumor purity and copy number, to estimate the number of clonal mutations in each sample. Mutations with a CCF > 0.9 were considered clonal. This accounted for 2243 mutations, 1963 (87.5%) of which were missense mutations (Additional file 2). We observed that clonal mutations were overrepresented among recurrently mutated genes (see previous section): 73 out of 137 mutations in these genes were clonal. This is significantly higher than the expected number given that 28% of the mutations were clonal (Chi-square test,  $p < 0.0001$ ).

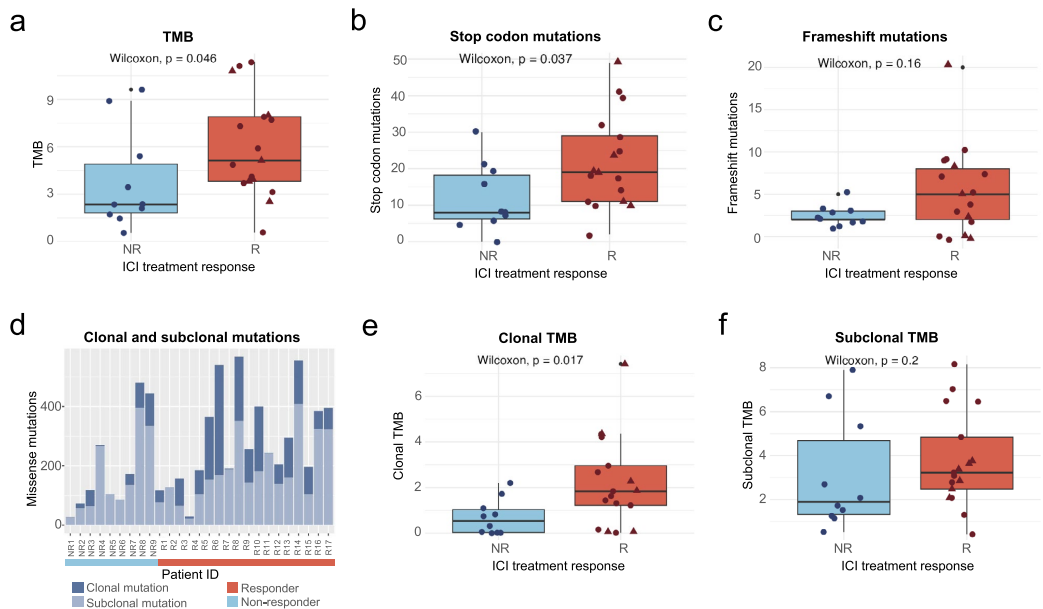
We found multiple lines of evidence that clonal mutations had a disproportionately higher impact in the response to the treatment when compared to other types of mutations. First, the samples of responders were strongly enriched in clonal mutations when compared to those of non-responders; the average percentage of clonal mutations in responders was twice the percentage in non-responders (34.32% versus 17.34%, Fisher's exact test  $p$  value = 0.007) (Fig. 2d). Second, when computing clonal TMB instead of total TMB, the differences between responders and non-responders clearly increased; for clonal mutations, the difference between the median values of the two groups showed an  $\times 1.7$  increase, and there was less overlap between the two distributions (Fig. 2e versus 2a). In contrast, there were no statistically significant differences in subclonal TMB between responders and non-responders (Fig. 2f, Wilcoxon test  $p$  value = 0.24). Third, we found that, for clonal mutations, there was an enrichment in APOBEC induced mutations in responders compared to non-responders ( $p$  value = 0.03, Figure S2); this effect was not observed when considering all non-synonymous mutations ( $p$  value = 0.44).

We checked if the above observations could be driven by the subset of complete responders. However, we did not observe a statistically significant difference in total, clonal or subclonal TMB when comparing complete responders and partial responders (Figure S3), indicating that this is not the case.

To better understand the possible implications for the prediction of response to ICI, we applied different TMB thresholds in steps of 1 and we classified the patients into responders (observed TMB above threshold) and non-responders (observed TMB below threshold). Then, we calculated the true positive and true negative rate for each threshold. Clonal TMB performed best in the threshold model with an AUC of 0.77, followed by total TMB with an AUC of 0.72 and subclonal TMB with an AUC of 0.62 (Fig. 3). Clonal TMB was particularly useful to identify true responders while keeping the false positive rate low (specificity 0.8–0.9).

Inspection of available mutation data from an independent urothelial cancer cohort, published by Snyder et al., also indicated that mutations occurring at high frequencies were especially important in explaining the response<sup>8</sup>. As no information on tumor cell purity or cell fraction was available for this dataset, we used the variant allele frequency (VAF), instead of the cancer cell fraction (CCF), as a proxy of mutation frequency. Similar to what we observed in our cohort, the difference between the median TMB values of responders to non-responders increased with increasing VAF (Fig. 4). The maximum difference was at VAF > 0.45, which essentially represents clonal mutations.

**The positive relationship between predicted neoantigens and response mirrors that observed for TMB.** One likely explanation for the association of TMB with response to treatment is that a subset of



**Figure 2.** TMB and clonal TMB are significantly associated with the response to ICI. (a) Relationship between TMB and response to ICI treatment. TMB values were significantly higher in responders than in non-responders (Wilcoxon test,  $p$  value = 0.046). (b) Relationship between stop mutations and response to ICI treatment. Nonsense mutations were significantly higher in responders than in non-responders (Wilcoxon test,  $p$  value = 0.037). (c) Relationship between frameshift mutations and response to ICI treatment. Frameshift mutations showed a trend of being more abundant in responders than non-responders but the difference was not statistically significant ( $p = 0.3$ ). (d) Number of mutations per patient divided in clonal and subclonal. The proportion of clonal mutations is significantly higher in responders (Fisher's exact test  $p$  value = 0.007). (e) Relationship between clonal TMB and response to ICI treatment. Clonal TMB values were significantly higher in responders than in non-responders (Wilcoxon test,  $p$  value = 0.017). (f) Relationship between subclonal TMB and response to ICI treatment. Responders tend to have higher values but the difference between responders and non-responders is not statistically significant. NR no responders, R responders, triangle shape represents complete responders among the group of responders.

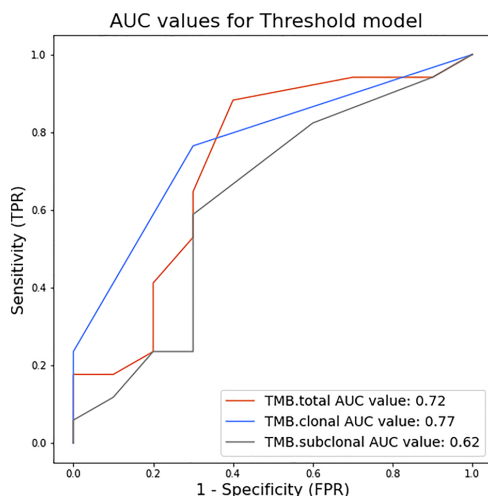
these mutations will generate neoantigens that trigger an immune system response against the tumor. Among neoantigen presentation, binding affinity is thought to be the most selective step<sup>2</sup>. We predicted the MHC I binding affinity of all possible peptides originating from missense mutations in the tumors using the software NetMHCpan 4.0<sup>31</sup>. Because MHC I receptors are highly variable among individuals, and each molecule has different epitope affinities<sup>32</sup>, we first computed the human leukocyte antigen (HLA) genotype for each patient and then used the specific HLA subtypes for the predictions. Following the NetMHCpan recommendations, we defined the set of putative MHC I bound peptides as those within the 2% top binding rank.

Regarding the relationship with ICI treatment response, we observed significantly more binders in responders than in non-responders ( $p$  value = 0.035, Fig. 5a). The difference became more significant when only considering clonal mutations ( $p$  value = 0.015, Fig. 5b). In contrast, no significant difference between treatment groups was found in the case of subclonal mutations ( $p$  value = 0.13, Fig. 5c). Similar conclusions were drawn when using the concentration that inhibits 50% binding of the fluorescein-labeled reference peptide ( $IC_{50}$ ), or more stringent thresholds (Figure S4), as well as when using a second MHC binding prediction software, MHCflurry 2.0 (Figure S5). The results are consistent with a positive effect of the number of neoantigens that are being presented on the response to the treatment.

Factors other than the peptide MHC binding affinity can also influence neoantigen presentation and immunogenicity<sup>2</sup>. One such factor is the stability of the peptide-MHC (pMHC) complex<sup>33,34</sup>. However, we did not find any significant difference between the number of predicted highly stable pMHC complexes and the response to treatment (Figure S6). Another approach to predict a peptide's relative immunogenicity is its differential agretopicity index (DAI)<sup>35</sup>. This index measures the difference in the binding affinity of the mutated peptide compared to its non-mutated counterpart. We observed that mutated peptides in non-responders tended to show lower DAI values compared to those in responders but, again, this trend did not achieve statistical significance (Figure S6).

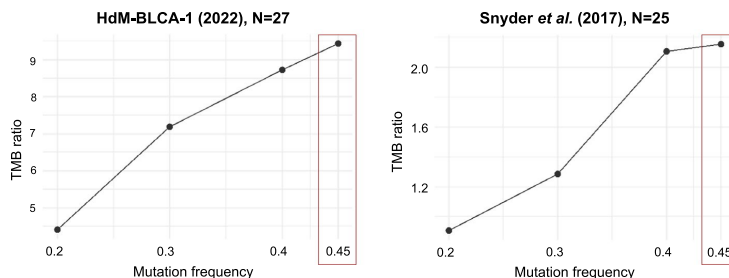
We then asked whether the difference between responders and non-responders could increase in the case of mutations that generated new binders. Using a binding rank < 2% threshold, we identified a total of 2175 new





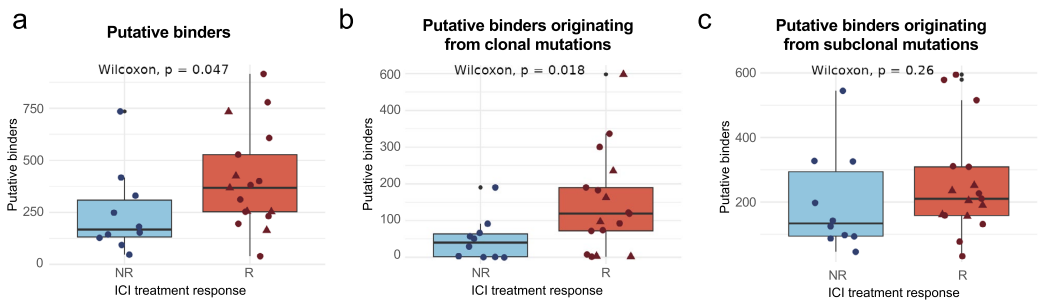
**Figure 3.** Clonal TMB performs better in separating the two response groups in a threshold model than total TMB. Thresholds to separate response groups in steps of TMB = 1 mut/mB were applied. Sensitivity or true positive rate (TPR) was calculated as the number of true positives divided by the number of true positives + false negatives. Specificity or false positive rate (FPR) was calculated as the number of true negatives divided by the number of true negatives plus false positives.

#### TMB ratio of responders to non-responders for different mutation frequencies



**Figure 4.** Effect of mutation frequency on the discrimination between responders and non-responders. TMB ratio: ratio of the median TMB of responders versus the median TMB of non-responders. Mutation frequency: minimum variant allele frequency that we consider to calculate TMB. The difference in TMB between responders and non-responders increases with mutation frequency, with maximum values at a mutation frequency of 0.45 in both cohorts.

putative peptide binders created by cancer-specific mutations, and another 1901 peptides in which the mutation caused the peptide to lose its ability to bind to MHC I. The new binders were enriched in hydrophobic amino acids, including tyrosine (Y), phenylalanine (F), leucine (L) and tryptophan (W), as well as histidine (H), when compared to mutations causing no change in binder status (Figure S7). In contrast, mutations associated with loss of binding to MHC I were enriched in cysteine (C) and, to a lower extent, glycine (G). We also observed that amino acid replacements that changed the binding status of the peptide tended to be located in the 2nd and 9th position of the peptide. These results are consistent with previous observations that hydrophobic residues are associated with increased immunogenicity<sup>36,37</sup>, and that the 2nd and 9th amino acids of the peptide are anchor positions to the MHC I receptor<sup>37,38</sup>. When responders and non-responders were compared, we again observed increased separation between the groups for clonal mutations compared to other types of mutations (Figure S8). The magnitude of this effect, however, was not larger than when all possible kinds of neoantigens were considered.



**Figure 5.** The positive relationship between predicted neoantigens and response mirrors that observed for TMB. (a) Relationship between the number of putative binders and the response to the treatment. Responders have a significantly higher number of putative binders than non-responders (Wilcoxon test,  $p$  value = 0.047). (b) Relationship between putative binders originating from clonal mutations and response to ICI treatment. Number of putative binders originating from clonal mutations is significantly higher in responders than in non-responders (Wilcoxon test,  $p$  value = 0.018). (c) Relationship between putative binders originating from subclonal mutations and response to ICI treatment. No significant difference can be observed for the number of putative binders originating from subclonal mutations between responders and non-responders. NR no responders, R responders, triangle shape represents complete responders among the group of responders.

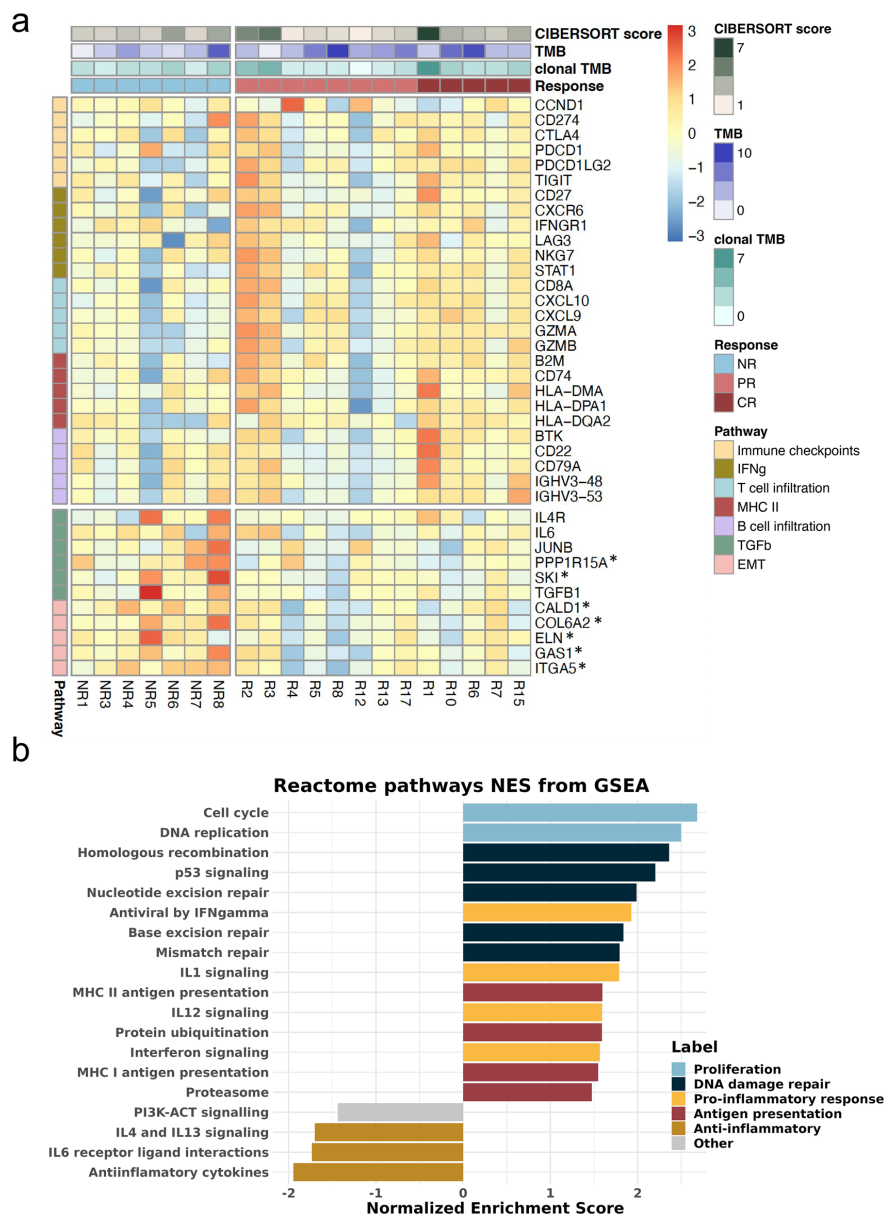
We also examined whether, independently of the response, the loss of binder mutations was relatively more frequent among clonal mutations than subclonal ones, this could be expected if this type of mutation is particularly favored in the initial stages of cancer when mechanisms to avoid immune system surveillance might be weaker. We found that, while the ratio of the number of loss of binder mutations versus gain of binder mutations was indeed higher for clonal than for subclonal mutations (0.93 versus 0.84, respectively), the difference was not statistically significant.

**Pathways associated with response and immune response markers.** In addition to the aforementioned genomic characteristics of neoantigens and the peptide-MHC complex, the tumor environment and other molecular mechanisms are known to play a crucial role in the activation of an immune response. We used the gene expression data to impute tumor-immune cell infiltration abundances using CIBERSORT. We found no significant differences in the overall immune infiltration score between responders and non-responders, but the former group had a significantly higher fraction of CD4 memory-activated T cells ( $p = 0.029$ , Figure S9a). Next, we examined the gene expression patterns of different immune markers and gene expression signatures (RNA-Seq data for 20 patients, 13 responders, 7 non-responders). We found that the median expression value of pro-inflammatory markers, immune checkpoints and MHCII antigen presentation genes, tended to be higher in responders than non-responders (Fig. 6a, Figure S9b). In their multivariable predictor model, Litchfield et al.<sup>18</sup> reported gene expression values of cytokine *CXCL9* to be among the strongest predictors for ICI response. In our cohort, *CXCL9* had more than twice the median value in the responders group (Figure S9b). Other markers related to CD8 T cell immune response, such as *CD45*, *CD8A*, and interferon-gamma pathway genes showed a similar expression pattern. Genes involved in B cell-mediated immunity and MHC class II also tended to be enriched in responders. On the contrary, gene expression of the transcription growth factor beta (*TGF- $\beta$* ) was found to be decreased in responders compared to non-responders (Fig. 6a). This result is expected given that *TGF- $\beta$*  is generally associated with an immunosuppressive effect<sup>39</sup>.

Analyzing the expression patterns of combinations of genes instead of individual genes can sometimes provide a clearer signal. Using gene set enrichment analysis (GSEA) identified several pathways significantly enriched in the responders group (FDR  $p$  value < 0.05). This included DNA damage repair (DDR), proliferation, apoptosis, ubiquitination and pro-inflammatory functions upregulated in responders (Fig. 6b, Figure S10). These pathways are very interrelated and are expected to play a role in ICI response. Additionally, non-responders were characterized by higher expression of anti-inflammatory cytokines like *IL4* and *IL6*, extracellular matrix organization and *TGF- $\beta$*  related pathways (Fig. 6b, Figure S10). Analyses performed grouping the genes by Gene Ontology (GO) classes or using molecular pathways produced consistent results (Figures S10, S11a).

We then investigated if there were any differences in gene expression patterns between complete responders ( $n = 5$ ) and partial responders ( $n = 8$ ). We found significant differences at the level of T cell-mediated and humoral immunity (Figure S11). Therefore, immune players beyond the classical MHC-I-CD8 T axis, such as those mediated by MHC class II, might be important.

Another interesting observation was that the subset of responders that had the lowest number of mutations (samples R1, R2 and R3; TMB < 3.5 mut/Mb) also had the highest levels of T and/or B cell infiltration markers (Fig. 6b). These patients would most likely have been missed by any TMB threshold method (the majority of the responders had TMB > 5 mut/Mb) but could have been identified using a signature based on immune markers.



**Figure 6.** (a) Responders with low clonal TMB show expression of genes connected to immune infiltration, while non-responders have higher expression of immune suppression. Gene expression for selected marker genes sorted by pathways. Expression values are normalized and log2cpm transformed. The heatmap is scaled by row. Annotation bars show immune infiltration as CIBERSORT score, tumor mutational burden (TMB) and clonal tumor mutational burden (clonal TMB) and response to ICI treatment. Columns indicate the patient tumor samples. NR no responder, PR partial responder, CR complete responder. Only a few genes were statistically significant in the DE analysis (\*p value < 0.05). (b) REACTOME pathways connected with proliferation, DNA damage repair, antigen presentation and pro-inflammation are significantly enriched in responders, and anti-inflammation pathways are enriched in non-responders. Normalized enrichment score for selected pathways significantly related to response obtained from GSEA analysis (adjusted  $P < 0.05$ ; comparing 13 responders and 7 non-responders). The complete list of pathways with the included genes is provided in Additional file 3.

## Discussion

Our findings indicate that the number of clonal non-synonymous mutations (clonal TMB) is the strongest predictor of the response to ICI in advanced urothelial cancer, suggesting that focusing on this biomarker could significantly improve the identification of the patients who are more likely to benefit from treatment. The results are well-aligned with previous findings for melanoma and non-small cell lung cancer, in which the burden of clonal neoantigens and a low intra-tumor heterogeneity have been associated with an increased response to ICI<sup>15,16</sup>. In the study of Miao et al., which included a range of cancer types including urothelial cancer, the authors reported that the number of clonal non-synonymous mutations was significantly higher in complete responders than in partial responders or non-responders<sup>17</sup>. Instead, we could not see differences in the number of clonal mutations between complete responders and partial responders. Finally, Litchfield et al. (2021) constructed multivariate predictive models and observed that models considering only TMB and *CXCL9* attained AUC values between 0.63 and 0.79 depending on the test cohort, which was only marginally lower than the full model containing eleven different variables<sup>18</sup>.

We found that the separation between responders and non-responders in our cohort increased as we considered mutations with a higher allele frequency, the same effect being observed using data from an independent urothelial cancer cohort by Snyder et al.<sup>8</sup>. Using a mutation threshold predictive model, the AUC values for clonal TMB were higher than those for total TMB (0.77 versus 0.72). In addition, we found that the ratio of clonal versus subclonal mutations was significantly higher in responders than in non-responders, suggesting that tumors in the responder group are less heterogeneous. A study using a UV-induced mouse melanoma model showed that tumor heterogeneity diminishes the immune response<sup>16</sup>. Taken together, the results indicate that the predictive power of TMB can be significantly increased if we only consider clonal mutations.

The APOBEC signature has been previously associated with the response to ICI in different types of cancer including urothelial cancer<sup>17,38</sup>. APOBEC-mediated mutagenesis becomes activated in early disease stages<sup>21,24,40</sup> and has been associated with mutation hotspots of several cancer driver genes<sup>41</sup>. In urothelial cancer, the APOBEC signatures are suggested to drive high TMB by introducing genomic instability<sup>42</sup>. Additionally, both the APOBEC signature and high TMB are related to better survival in urothelial cancer patients, independently of treatment and in the context of anti-PD-L1/PD-1 ICI<sup>42,43</sup>. In our study, we did not find a significant positive association between APOBEC and the response to treatment, except when we measured the APOBEC signature specifically in clonal mutations. A potential explanation could be that APOBEC mutations are associated with increased peptide hydrophobicity, as shown by Boichard et al.<sup>38</sup>. This could increase the likelihood of generating immunogenic neoantigens<sup>36,37</sup>.

A detailed inspection of the mutations revealed that many known oncogenes and tumor suppressor genes were frequently mutated in our patient cohort. The topmost mutated genes were *ELF3*, *KMT2D*, *ARID1A*, *TP53* and *PRKDC*; these genes were mutated in 6 patients or more (> 20% of the patients). Interestingly, a recent study found that recurrent somatic mutations in several cancer-related genes and pathways, including the MAPK signaling and TP53 cycle pathways, increased the predictive power of the predictive models over just using TMB. Given that a large portion of the mutations in these genes are clonal, these results also point to the importance of tumor clonality in the response to immunotherapy.

We also examined the relationship between the number of predicted MHC-I binding peptides per patient and ICI treatment response as described in previous studies<sup>31,44</sup>. We found that, similar to TMB, the clonality of the mutations generating the predicted neoantigens increased the difference between response groups. For a subset of the somatic mutations, the mutation increased the binding of the peptide to MHC I above the threshold, potentially generating a new binder. Again, we found that the number of new binders was significantly higher in responders versus non-responder in the case of clonal mutations. Further studies based on experimental data might help disentangle the importance of different types of neoantigens. It has been reported that mutations are more likely to be observed in tumors if the resulting peptides have low affinity for the patient's MHC I receptors<sup>45</sup>. Although we observed a tendency for increased loss of binding mutations in clonal mutations versus subclonal ones, the results were not conclusive. We expect, however, that these analyses will encourage future studies to understand selective processes involved in immune evasion.

The gene expression analysis of immune markers and immune cell invasion indicated that T cell immune infiltration is an important factor in determining response to the treatment, as found in other studies<sup>8,18,39</sup>. A high neoantigen load will not elicit an immune response if the tumor is not invaded by T cells and consequently, the ICI treatment will be ineffective. Using gene set enrichment we observed that genes involved in proliferation, DNA damage response, antigen presentation, and pro-inflammatory responses were significantly associated with response to treatment. Additionally, we saw higher immune infiltration of CD4 memory T cells as imputed by deconvolution, and increased expression of CD8 T cell markers in responders. High tumor infiltration of CD8 T cells has been associated with improved clinical outcome through the activation by MHC I presented antigens in urothelial bladder cancer<sup>46</sup>. On the contrary, T regulatory (Treg) cells are described to be tumor-promoting agents<sup>47</sup>. Our results indicated a negative association between ICI response and both Treg and Th cells, although it did not achieve statistical significance. The latter type of T cells expresses a large number of PD-1 on their surface. It has been suggested that anti-PD-1 treatment could cause immune-related adverse events (irAEs) by hyperactivation of Th, provoking autoimmunity<sup>48</sup>. High levels of Th cells have been recently correlated to irAEs in urothelial bladder cancer patients treated with PD-L1<sup>49</sup>. The role of B cells in ICI therapy in urothelial cancer has been sparsely studied, despite their high expression of PD-L1<sup>50</sup>. While a positive correlation between B cells and improved survival was suggested for different cancer types<sup>51,52</sup>, other studies reported B cell content not to be associated with response to anti-PD-1 in melanoma<sup>53</sup>. Our results indicate that B cell infiltration levels are quite heterogeneous among responders. We further observed that expression of TGF- $\beta$

tended to be lower in responders than non-responders. The TGF- $\beta$  signature is associated with fibroblasts in cells excluded from the tumor parenchyma and has been previously associated with a lack of response and resistance to immunotherapy<sup>39</sup>.

Combining information from WES and RNASeq might help to better discriminate between responders and non-responders. In our set of responders, there were three patients with an abnormally low number of missense mutations. Interestingly, these three patients were also the ones with the highest immune infiltrate based on the gene expression analysis, indicating that, at least in these cases, a high number of immune cells was the key. However, the interplay between different variables might be quite intricate. The use of machine learning approaches to build predictive models could be a promising approach, but this requires a large amount of data, and the accuracy of the predictions is still relatively low. In linear regression models, the effects of different variables are additive, which hampers the identification of different classes of responders. Models capable of detecting the biology underneath the data, and the dependencies between different variables, need to be developed.

Finally, a few important limitations should be mentioned. The pipeline focused on missense SNVs, without taking indels or splice-isoforms into account. Peptides occurring from such mutations are expected to be highly immunogenic and they could be contributing in a significant manner in explaining the response to immunotherapy<sup>24</sup>. However, they are much less frequent than missense mutations, and larger cohorts would be necessary to investigate their importance. Regarding the peptide-MHC complex analysis, we only considered MHC class-I-restricted neoantigens as, up to date, the prediction algorithms for MHC II are less reliable and with lower accuracy<sup>2</sup>, but MHC-II-mediated immune response might also be relevant in ICI response<sup>35,36</sup>. It is worth mentioning that antigen preprocessing steps, such as peptide cleavage, its affinity to TAP protein, or ERAP trimming, are also important to predict neoantigens. To account for peptide processing, we also performed searches with MHCflurry 2.0, reaching similar conclusions as those obtained with NetMHCpan 4.0<sup>2,31,57</sup>. Finally, the bulk nature of our data (both WES and RNASeq) restricts all of our results to in-silico estimations. Future studies using single cell profiling will be needed to confirm our findings related to clonality and cell-type abundances. All neoantigen properties were based on predictions, which poses important limitations to the interpretation of the results. While performing immune-peptidomics experiments would provide a more realistic view of the neoantigen landscape, these experiments are not possible with FFPE samples.

This study presents a comprehensive and systematic analysis of a set of different biomarkers and their role as predictive biomarkers of response to ICI treatment in advanced urothelial cancer patients. We provide evidence that clonal TMB is a stronger predictor of response to ICI than total TMB. The results also suggest that, in some cases, patients with low TMB can respond to the treatment; this is generally associated with high levels of immune markers. Our study provides new data supporting that more homogeneous cancers in terms of TMB might be more likely to respond to ICI, and suggest that non-additive effects of different variables should be considered in future efforts to develop predictive models. Building predictive models able to combine different possible responder profiles might help address some of the present challenges for the management of advanced urothelial cancer with ICI.

## Material and methods

**Patient data.** The data analyzed here was derived from biological samples of 27 metastatic urothelial cancer patients (4 female and 23 male patients) treated in Hospital del Mar with anti-PD-1/PD-L1 ICIs. This study was approved by the Institutional Review Board of Hospital del Mar. For 27 patients, whole exome data was obtained from tumors and blood samples. Patients were classified as having clinical benefit (R—responders) if they had a partial or complete response in tumor burden, or having no clinical benefit (NR—non-responders) if they had progressive disease as the best response to treatment. Radiologic responses or progressive disease were defined as per the Response Evaluation Criteria In Solid Tumors (RECIST) criteria 1.1. Of the 27 patients, 17 were responders to the ICI treatment with 5 being complete responders. DNA and RNA extraction from formalin-fixed paraffin-embedded (FFPE) tumor specimens and blood samples was performed according to our experience<sup>58,59</sup>. Whole exome sequencing (WES) and RNA-sequencing (RNASeq) were done by the Centro Nacional de Análisis Genómico, Barcelona (CNAG). Sample coverage was analyzed by qualimap (Version 2.2.1) with default parameters and copy number analysis was conducted using ControlFreeC (Version 5.6) using the BAF options.

**Variant calling.** The preprocessing of the raw sequencing data was conducted at CNAG following the best practice GATK4 pipeline (version 4.0.8). Sequence reads were mapped to the reference genome (hs37d5) using BWA (version 0.7.17) to obtain SAM/BAM files sorted by coordinates. To mitigate biases introduced by data generation steps such as PCR amplification, duplicates were marked, and base quality scores were re-calibrated, as variant calling algorithms rely heavily on the quality scores assigned to the individual base calls in each sequencing read. Mutect2 (version 40.1.2) and Strelka2 (version 2.9.10) were used as variant callers comparing somatic and germline samples for each patient. Mutations were annotated using VEP (Version 104). SNVs that pass the default filters of both, Mutect2 and Strelka2, were further filtered for population-wide allele frequency under 5% (gnomAD), a minimum sample depth of 30, and a minimum alternative allele depth of 3. Further analyses were performed focusing on missense mutations to facilitate the comparison of mutated and non-mutated peptides. TMB was measured using the function `tmb()` of the `maftools` R-package (version 2.10.0)<sup>22</sup>, estimating the number of non-synonymous mutations per capture size of 50 megabases as this is the target region of the kit used. The TMB threshold model was built by applying different thresholds in steps of 1 to the dataset, separating the patients into responders (TMB above threshold) and non-responders (TMB below threshold). From there, the number of correctly and misclassified patients was used to obtain specificity (True

positive rate = true positives / (true positives + false negatives) and sensitivity (1 – True negative rate = 1 – (true negatives / (true negatives + false positives))) for each threshold.

**APOBEC enrichment estimation.** To assess the apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) mutational signature enrichment, the trinucleotideMatrix function of the maftools R-package was used<sup>22</sup>. The function compares the enrichment of C>T mutations occurring in TCW motives over the total of C>T mutations in the given sample to a background of occurring cytosines and TCW motives. Samples with an enrichment score > 2 and p value < 0.05 were considered significant.

**Clonality.** Clonality was defined as mutations with a cancer cell fraction (CCF) above 0.9 and it is calculated as:

$$CCF = VAF/p * (2 * (1 - p) + c * p)$$

c = copy number, p = purity.

Sample-specific tumor purity and copy number are included to adjust for the variable tumor content and copy number changes following Tarabichi et al.<sup>60</sup>. Clonal mutations were therefore defined as SNVs with a CCF > 0.9. Tumor purity and local allele-specific copy number were computed using ASCAT (version 3.0). We used default parameters to run ASCAT, except for gamma = 1 as recommended by the developers for WES data<sup>61</sup>. The input for ASCAT was generated using alleleCount (version 4.3.0). The loci file to obtain allele-specific copy numbers was downloaded from the nf-core/sarek pipeline (release 3.1.1)<sup>62</sup>. We excluded all loci not covered by the exome target regions from the loci file by building the intersect of loci and BED file that was used for the WES data.

**Neoantigen and HLA prediction.** For each patient, the according 4-digits HLA genotype was determined following the nf-core/hlatyping pipeline (release 1.2.0)<sup>63</sup> using the HLA genotyping algorithm OptiType (version 1.3.5)<sup>32</sup>. We used blood samples to type each patient's HLA alleles running OptiType at default parameters. Secondly, we computed all possible 9-mer peptides that encompass a mutation, as well as their non-mutated counterpart. This was done using an in-house python script with a sliding window. The peptide sequences were downloaded from Ensembl GRCh37, release 75.

**Binding affinity.** To predict the binding affinity of tumor and germline peptides to MHC-I molecules, NetMHCpan (version 4.0) and MHCflurry (version 2.0.4) were used<sup>31,57,64</sup>. By default, NetMHCpan 4.0 labels peptides with a binding rank under 2% as weak binder (WB) and under 0.5% as a strong binder (SB), as defined by the program. The rest of the peptides were predicted to have no binders. We additionally applied a threshold using the predicted concentration that inhibits 50% binding of the fluorescein-labelled reference peptide (IC<sub>50</sub>). Peptides with a binding affinity IC<sub>50</sub> < 500 nM were labelled to be weak binding and peptides with IC<sub>50</sub> < 50 nM being strong binding.

By comparing the classification of the mutated peptides and their non-mutated counterparts, we identified cases in which the mutation was predicted to cause a peptide to become a new binder and cases in which the opposite happened (no binder to WB, and WB to no binder, respectively). We compared the frequency of all possible amino acid replacements in these peptides to those occurring in peptides for which no change in binding status was predicted (WB to WB, or no binder to binder). The latter provided an expectation against which to compare the observations. This allowed to identify which amino acids were most strongly associated with the gain of new binders or the loss of existing binders. We performed similar analyses in relation to the peptide position in which the change was observed.

**Binding stability.** Binding stability was predicted using netMHCstabpan (version 1.0)<sup>34</sup>. The applied threshold for long binders was 1.4 h following the approach of Wells et al.<sup>65</sup>.

**Agretopicity.** To rank the mutated peptides by improved MHC-I binding affinity compared to their wildtype counterparts, we calculated the differential agretopicity index (DAI) as described by Duan et al.<sup>35</sup> and used in previous studies<sup>18,66</sup>. Following the approach of Rech et al.<sup>44</sup>, a threshold was applied at DAI > 9 for neoantigens with high differential binding affinity compared to their non-mutated counterparts. The threshold for peptides with high DAI was set to 9 as this included only mutated peptides with a binding affinity of more than > 50 nM.

**Gene expression analysis.** For 22 of the 27 patients, RNASeq data was generated. Raw sequencing reads were mapped with STAR (version 2.6.0)<sup>67</sup>. Gencode (release 29) based on the GRCh38.p13 reference genome and the corresponding gene transfer format (GTF) file was used. The table of counts was obtained with FeatureCounts function in the package subread (version 1.6.4)<sup>68</sup> with the previously mentioned GTF file. Genes having less than 10 counts in at least 7 samples were excluded from the analysis. Raw library size differences between samples were treated with the weighted “trimmed mean method” TMM<sup>69</sup> implemented in the edgeR package (version 3.36.0)<sup>70</sup>. The normalized log<sub>2</sub>CPMs were used in order to make hierarchical clustering and PCA to assess batch effects and outliers. One sample (R16) was removed from further analysis (Figure S12). For the differential expression (DE) analysis, we used the DGEList object with TMM normalization factors as input for the voom approach of the limma package (version 3.50.0), which models the mean–variance relationship of the log-counts with precision weights. The results of the DE analysis are presented as volcano plots in the supplementary figure S13. We assessed the incursion of surrogate variables and covariates but deemed it unnecessary since the results did not improve. Pre-Ranked Gene Set Enrichment Analysis (GSEA)<sup>71</sup> implemented in clusterProfiler<sup>72</sup>



package (version 3.18.0) was used in order to retrieve enriched functional pathways. The ranked list of genes was generated using the  $-\log(p.val) * \text{signFC}$  for each gene from the statistics obtained in the DE analysis with limma. Functional annotation was obtained based on the enrichment of gene sets belonging to gene set collections in Molecular Signatures Database (MSigDB, version 7.5). The complete results are provided in Additional data file 3. The association of previously described gene signatures with response was tested using the normalized log2CPM with the GSVA package<sup>73</sup>. TPM values were used for the deconvolution with CIBERSORTx<sup>74</sup> method with the LM22 gene signature and using B-mode batch correction and absolute mode. The CIBERSORT estimated abundances and the gene signatures scores computed with GSVA are detailed in Additional File 2 and 3, respectively.

**Application of clonality threshold to an independent dataset.** As an independent dataset, we analyzed the publicly available mutations of 26 urothelial cancer patients treated with atezolizumab (an anti-PD-L1 antibody) previously published by Snyder et al.<sup>8</sup>. The downloaded somatic mutations were reannotated using VEP (Version 104). Due to the low number of mutations called by Strelka2 (average of 25 mutations/patient, range 11–62) it was not possible to build the intersect of the two callers, Strelka2 and Mutect2, as it was done for the dataset of Hospital del Mar. Instead, the union was built of both lists, following the method described by Snyder et al. Mutations were then filtered as described above (gnomAD < 5%, sample depth > 30X, alternative allele depth > 3X). Following Snyder et al., patient 4072 was excluded due to low coverage. Clinical treatment response was evaluated as described above. No information on tumor sample purity and cell fraction were available. Thus, the effect of clonality was estimated by applying different thresholds of minimum tumor variant allele frequency (VAF) from 0.1 to 0.7. Above 0.7, only one responding patient was found to have mutations passing the threshold. The TMB ratio was calculated as TMB in responders divided by TMB in non-responders. To compare these results, the same calculation was repeated for the Hospital del Mar dataset.

**Statistical tests and graphs.** All plots included in this manuscript were generated using R (version 4.1.2) and RStudio (version 1.4.1106). For data integration we also used scripts written in python (version 3.5.2). Wilcoxon signed-rank test was used to obtain the significant differences between treatment response groups.

**Ethical approval.** The presented study was approved by the institutional Ethics Committee for Clinical Investigation of the Hospital del Mar-IMIM, Barcelona, Spain (2016/6767/I) and conducted in accordance with the principles set out in the World Medical Association guidelines (Seventh revision of the 249 Declaration of Helsinki, Fortaleza, Brazil, 2013) for human subjects involved in medical investigations. All patients signed informed consent for the analysis of tumor biopsies for research purposes and biomarker assessment.

### Data availability

The dataset generated and analyzed during the current study is available in the EGA repository under the study ID EGAS00001007086.

Received: 5 April 2023; Accepted: 11 September 2023

Published online: 15 September 2023

### References

- Antoni, S. et al. Bladder cancer incidence and mortality: A global overview and recent trends. *Eur. Urol.* **71**, 96–108 (2017).
- Richters, M. M. et al. Best practices for bioinformatic characterization of neoantigens for clinical utility. *Genome Med.* **11**, 1–21 (2019).
- Li, K. et al. PD-1 suppresses TCR-CD8 cooperativity during T-cell antigen recognition. *Nat. Commun.* **12**, 1–13 (2021).
- Inman, B. A. et al. PD-L1 (B7-H1) expression by urothelial carcinoma of the bladder and BCG-induced granulomata: Associations with localized stage progression. *Cancer* **109**, 1499–1505 (2007).
- Bellmunt, J. et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N. Engl. J. Med.* **376**, 1015–1026 (2017).
- Haslam, A. & Prasad, V. Estimation of the percentage of US patients with cancer who are eligible for and respond to checkpoint inhibitor immunotherapy drugs. *JAMA Netw. Open* **2**, 25 (2019).
- Lopez-Beltran, A. et al. Immune checkpoint inhibitors for the treatment of bladder cancer. *Cancers* **13**, 131 (2021).
- Snyder, A. et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. *PLoS Med.* **14**, 25 (2017).
- Lv, J., Zhu, Y., Ji, A., Zhang, Q. & Liao, G. Mining TCGA database for tumor mutation burden and their clinical significance in bladder cancer. *Biosci. Rep.* **40**, 25 (2020).
- Graf, R. P. et al. Tumor mutational burden as a predictor of first-line immune checkpoint inhibitor versus carboplatin benefit in cisplatin-unfit patients with urothelial carcinoma. *JCO Precis. Oncol.* **6**, 25 (2022).
- McGrail, D. J. et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* **32**, 661–672 (2021).
- Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
- Sha, D. et al. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* **10**, 1808–1825 (2020).
- Stronen, E. et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
- McGranahan, N. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **1979**(351), 1463–1469 (2016).
- Wolf, Y. et al. UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell* **179**, 219–235.e21 (2019).
- Miao, D. et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat. Genet.* **50**, 1271–1281 (2018).
- Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).

19. Robertson, A. G. *et al.* Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017).
20. Kim, J. *et al.* Invasive bladder cancer: Genomic insights and therapeutic promise. *Clin. Cancer Res.* **21**, 4514–4524 (2015).
21. Roberts, S. A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
22. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
23. Weinstein, J. N. *et al.* Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
24. Nordtöft, I. *et al.* Mutational context and diverse clonal development in early and late bladder cancer. *Cell. Rep.* **7**, 1649–1663 (2014).
25. Böck, M. *et al.* Identification of ELF3 as an early transcriptional regulator of human urothelium. *Dev. Biol.* **386**, 321–330 (2014).
26. Kandath, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
27. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875–878 (2011).
28. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* (1979) **362**, 3593 (2018).
29. Gajic, Z. Z., Deshpande, A., Legut, M., Imieliński, M. & Sanjana, N. E. Recurrent somatic mutations as predictors of immunotherapy response. *Nat. Commun.* **13**, 1–11 (2022).
30. Fehrenbacher, L. *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837–1846 (2016).
31. Jurtz, V. *et al.* NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 25 (2017).
32. Szolek, A. *et al.* OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
33. Harndahl, M. *et al.* Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur. J. Immunol.* **42**, 1405–1416 (2012).
34. Rasmussen, M. *et al.* Pan-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol.* **197**, 1517–1524 (2016).
35. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
36. Chowell, D. *et al.* TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes. *Proc. Natl. Acad. Sci. USA* **112**, E1754–E1762 (2015).
37. Capietto, A.-H.H. *et al.* Mutation position is an important determinant for predicting cancer neoantigens. *J. Exp. Med.* **217**, 25 (2020).
38. Boichard, A. *et al.* APOBEC-related mutagenesis and neo-peptide hydrophobicity: Implications for response to immunotherapy. *Oncimmunology* **8**, 25 (2019).
39. Mariathasan, S. *et al.* TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
40. Bellmunt, J. *et al.* Genomic predictors of good outcome, recurrence, or progression in high-grade T1 non-muscle-invasive bladder cancer. *Cancer Res.* **80**, 4476–4486 (2020).
41. Shi, M. J. *et al.* Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. *Genome Med.* **12**, 25 (2020).
42. Glaser, A. P. *et al.* APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget* **9**, 4537 (2018).
43. Galsky, M. D. *et al.* Nivolumab in patients with advanced platinum-resistant urothelial carcinoma: Efficacy, safety, and biomarker analyses with follow-up from CheckMate 275. *Clin. Cancer Res.* **26**, 5120–5128 (2020).
44. Rech, A. J. *et al.* Tumor immunity and survival as a function of alternative neoepitopes in human cancer. *Cancer Immunol. Res.* **6**, 276 (2018).
45. Marty, R. *et al.* MHC-I genotype restricts the oncogenic mutational landscape. *Cell* **171**, 1272–1283.e15 (2017).
46. Sharma, P. *et al.* CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc. Natl. Acad. Sci. USA* **104**, 3967 (2007).
47. Kitamura, T., Qian, B. Z. & Pollard, J. W. Immune cell promotion of metastasis. *Nat. Rev. Immunol.* **15**, 73–86 (2015).
48. Gutiérrez-Melo, N. & Baumjohann, D. T follicular helper cells in cancer. *Trends Cancer* **2**, 5 (2023).
49. Herati, R. S. *et al.* PD-1 directed immunotherapy alters Tfh and humoral immune responses to seasonal influenza vaccine. *Nat. Immunol.* **23**, 1183 (2022).
50. Kim, S. S. *et al.* Role of B cells in responses to checkpoint blockade immunotherapy and overall survival of cancer patients. *Clin. Cancer Res.* **27**, 6075 (2021).
51. Schaafsma, E., Jiang, C. & Cheng, C. B cell infiltration is highly associated with prognosis and an immune-infiltrated tumor microenvironment in neuroblastoma. *J. Cancer Metastasis Treat.* **7**, 25 (2021).
52. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).
53. Damsky, W. *et al.* B cell depletion or absence does not impede anti-tumor activity of PD-1 inhibitors. *J. Immunother. Cancer* **7**, 1–7 (2019).
54. Turajlic, S. *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: A pan-cancer analysis. *Lancet Oncol.* **18**, 1009–1021 (2017).
55. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
56. Yi, R. *et al.* MHC-II signature correlates with anti-tumor immunity and predicts anti-PD-L1 response of bladder cancer. *Front. Cell Dev. Biol.* **10**, 25 (2022).
57. O'Donnell, T. J., Rubinstein, A. & Laserson, U. MHCflurry 2.0: Improved pan-allele prediction of mhc class i-presented peptides by incorporating antigen processing. *Cell Syst.* **11**, 42–487 (2020).
58. Esteve-Codina, A. *et al.* A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS One* <https://doi.org/10.1371/journal.pone.0170632> (2017).
59. Kim, Y. K., Han, H. S., Yoon, Y. S., Cho, J. Y. & Lee, W. Laparoscopic approach for right-sided intrahepatic duct stones: A comparative study of laparoscopic versus open treatment. *World J. Surg.* **39**, 1224–1230 (2015).
60. Tarabichi, M. *et al.* A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat. Methods* **18**, 144–155 (2021).
61. Van Looy, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
62. Garcia, M. *et al.* Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res* **9**, 63 (2020).
63. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278. <https://doi.org/10.1038/s41587-020-0439-x> (2020).
64. Reynisson, B. *et al.* NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
65. Wells, D. K. *et al.* Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834.e13 (2020).



66. Ghorani, E. *et al.* Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Ann. Oncol.* **29**, 271–279 (2018).
67. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
69. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 1–9 (2010).
70. 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).
71. 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).
72. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
73. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinform.* **14**, 1–15 (2013).
74. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

## Acknowledgements

We gratefully acknowledge the contribution of the patients included in the HdM-BLCA-1 cohort. Further, we would like to express our gratitude to Robert Castelo and Gabriel Santpere, as well as the scientists in their group, for their continuous support and advice throughout the research. This work has been supported by FIS PI16/00112 and FIS PI22/00171 (Spanish Health Ministry Grant “Fondo de Investigación Sanitaria”), as well as from Ayudas Fundación BBVA a Proyectos de Investigación Científica en Biomedicina 2021 and by Generalitat de Catalunya (2021SGR00776 and 2021SGR00042). LMB is funded by an INPhINIT PhD fellowship from “la Caixa” Foundation (ID 100010434), under the agreement LCF/BQ/DI21/11860060.

## Author contributions

J.B., A.R.V., M.M.A., J.P.B. designed and coordinated the study. J.B., A.R.V., N.J. and J.L. did the clinicopathological review of the cohort. A.R., O.A. and S.H.L. performed the experimental processing of samples. J.P.B., L.M.B., M.M.A. and S.V. performed the data analysis.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-42495-2>.

**Correspondence** and requests for materials should be addressed to M.M.A. or J.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023



## **3.2 Predicting immunotherapy response in advanced bladder cancer through a meta-analysis of six independent cohorts**

### **Full citation:**

Boll LM, Vázquez Montes de Oca S, Camarena ME, Castelo R, Bellmunt J, Perera-Bel J, Albà MM. Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts. *Nat Commun.* 2025 Feb 20;16(1):1213. doi: 10.1038/s41467-025-56462-0. PMID: 39979258; PMCID: PMC11842772.

**First authors:** Lilian Marie Boll and Sergio Vázquez Montes de Oca

**Status:** Published on February 20<sup>th</sup>, 2025

**Journal:** Nature Communications

**Correspondence:** malba@researchmar.net,  
jperera@researchmar.net, and joaquim\_bellmunt@dfci.harvard.edu

**Supplementary Material:** Appendix, section 5.2.

# Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts

Received: 27 April 2024

Accepted: 14 January 2025

Published online: 20 February 2025



Lilian Marie Boll<sup>1,5</sup>, Sergio Vázquez Montes de Oca<sup>1,5</sup>, Marta E. Camarena<sup>1</sup>, Robert Castelo<sup>2</sup>, Joaquim Bellmunt<sup>1,3</sup>✉, Júlia Perera-Bel<sup>1</sup>✉ & M. Mar Albà<sup>1,4</sup>✉

Advanced bladder cancer patients show very variable responses to immune checkpoint inhibitors (ICIs) and effective strategies to predict response are still lacking. Here we integrate mutation and gene expression data from 707 advanced bladder cancer patients treated with anti-PD-1/anti-PD-L1 to build highly accurate predictive models. We find that, in addition to tumor mutational burden (TMB), enrichment in the APOBEC mutational signature, and the abundance of pro-inflammatory macrophages, are major factors associated with the response. Paradoxically, patients with high immune infiltration do not show an overall better response. We show that this can be explained by the activation of immune suppressive mechanisms in a large portion of these patients. In the case of non-immune-infiltrated cancer subtypes, we uncover specific variables likely to be involved in the response. Our findings provide information for advancing precision medicine in patients with advanced bladder cancer treated with immunotherapy.

Immune checkpoint inhibitors (ICI) have been a major breakthrough in the treatment of advanced bladder cancer<sup>1,2</sup>. ICI can result in partial or even complete remission of the tumor, but it is only effective in a subset of the patients. Understanding which are the factors that underlie the response to ICI is key to improve our ability to predict response and to develop better therapeutic instruments<sup>3,4</sup>. While the number of somatic mutations in the tumor, or tumor mutational burden (TMB), is positively associated with ICI response, TMB alone has a moderate predictive power<sup>5,6</sup>. Another biomarker with conflicting results is the expression of the programmed cell death-ligand 1 (PD-L1) together with its receptor, programmed cell death protein 1 (PD-1), targeted by ICI drugs<sup>1,2,7,8</sup>. Immune players that have been associated with the response include the abundance of CD8<sup>+</sup> T cells, positively associated with the response in inflamed tumors, and the transforming growth factor  $\beta$  (TGF- $\beta$ ) signaling in fibroblasts, with a negative effect on the response to ICI<sup>9</sup>.

Bladder cancer (BLCA) shows a high degree of heterogeneity, and five muscle-invasive subtypes have been defined in the TCGA urothelial bladder carcinoma cohort<sup>10–12</sup>. The basal-squamous subtype is characterized by high expression of basal and stem-like markers, as well as immune markers. The luminal-papillary subtype corresponds to tumors with papillary morphology and enriched in *FGFR3* mutations. The luminal-infiltrated subtype shows low tumor purity and muscle-related signatures, and corresponds to the stroma-rich subtype in other classifiers<sup>11</sup>. The luminal subtype is associated with high levels of uroplakins UPK1A and UPK2. Finally, the neuronal class is associated with high levels of neuronal differentiation and development genes. Whether these different subtypes show a similar or different response rate to immunotherapy is a matter of debate. The luminal-infiltrated and basal-squamous subtypes have a high proportion of infiltrated immune cells and this has been hypothesized to result in more effective responses to ICI<sup>10,13</sup>. However, the results of a recent clinical trial

<sup>1</sup>Hospital del Mar Research Institute (HMRIB), Barcelona, Spain. <sup>2</sup>Department of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), Barcelona, Spain.

<sup>3</sup>Dana Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain.

<sup>5</sup>These authors contributed equally: Lilian Marie Boll, Sergio Vázquez Montes de Oca. ✉e-mail: [Joaquim\\_Bellmunt@dfci.harvard.edu](mailto:Joaquim_Bellmunt@dfci.harvard.edu); [jperera@researchmar.net](mailto:jperera@researchmar.net); [malba@researchmar.net](mailto:malba@researchmar.net)

have suggested that neuronal is the only subtype associated with a higher response<sup>14</sup>.

Increasing our ability to predict the response to ICI is key to providing better treatments to patients. This can be achieved by building computational predictive models capable of integrating disparate sources of omics data<sup>15–17</sup>. For these models to be robust, it is fundamental to gather data from a large number of patients. In order to address this challenge, we have assembled data from 707 tumors from six different cohorts of advanced BLCA patients treated with anti-PD-1/PD-L1. The models we have developed show high predictive accuracy and at the same time provide insights into the determinants of ICI response in the different subtypes.

## Results

### Building an advanced BLA meta-cohort

We extracted clinical and omics data for a total of 707 advanced BLCA patients treated with anti-PD-L1/PD-1 (Fig. 1). The patients were diagnosed with metastatic or locally advanced BLCA. Paired tumor-germline whole-exome sequencing (WES) data and tumor RNA sequencing (RNA-Seq) data were available for IMvigor210<sup>9</sup>, HdM-BLCA-1<sup>18</sup>, and SNY-2017<sup>5</sup>. For MIAO-2018<sup>5</sup>, WES data was collected. For the cohorts UC-GENOME<sup>19</sup> and UNC-108<sup>16</sup>, tumor RNA-Seq data, together with targeted DNA sequencing for a subset of patients, was available. To allow for a meaningful comparison of ICI therapy responders and non-responders, we considered the RECIST status of complete response (CR) and partial response (PR) as responders (class R) and PD (progressive disease) as non-responders (class NR). Stable disease (SD) and NE (not evaluable) were not considered for further analyses. We obtained a total of 466 patients classified as R or NR (163 and 303, respectively). For the majority of these patients ( $N = 348$ , 75%) both mutation and tumor gene expression data was available.

Investigation of clinical and demographic variables indicated that having a low functional status (ECOG score  $\geq 1$ ) and presence of liver metastasis were significantly associated with lack of response to ICI ( $\chi^2$ -test  $p$  value = 0.006 and  $p$  value = 0.001, respectively). Other characteristics such as sex, age, or smoker status were not significantly associated with the response (Supplementary Table 1).

### Analysis of mutational landscape

We obtained a high-resolution map of the most common somatic mutations in 121 previously defined BLCA-associated genes (Fig. 1, Supplementary Table 2). *TP53* was the most frequently mutated with 132 mutations followed by *KMT2D* (112 muts), *ARID1A* (78 muts), *ELF3* (67 muts), *KDM6A* (56 muts), and *FGFR3* (55 muts). These genes have also been found to be frequently mutated in BLCA tumors from TCGA<sup>10</sup> and other cancer cohorts<sup>20</sup>.

We tested if there was any association between mutations in these genes and the response to the treatment. Although mutations in the *FGFR3* are a hallmark for reduced immune infiltration<sup>10,11</sup>, we observed no relationship between *FGFR3* alterations and response to ICI, strengthening the conclusions of a previous report that was based on 103 patients, 17 of which had *FGFR3* alterations<sup>19</sup>. The only gene that was significantly associated with response was *ARHGEF12* (mutated in 11 responders and 1 non-responder, adjusted  $p$  value = 0.015). This gene encodes a Rho GTPase and its deletion has been associated with neuroblastoma differentiation and decreased stemness-related gene expression<sup>21</sup>.

### Missense and non-stop mutations are significantly associated with the response to ICI

We measured the tumor mutational burden (TMB), which is the number of non-synonymous mutations generated in the tumor, using WES data. For patients of the UNC-108 and UC-GENOME cohorts, we used the provided TMB values, which were derived from panel DNA data. To integrate the results of the different cohorts we transformed

the original TMB values to Z scores, separately for each cohort, and then combined the Z scores of the different cohorts ( $N = 378$ ).

We found that responders had significantly higher TMB than non-responders (Fig. 2A, B), consistent with previous findings<sup>3,9,15,22</sup>. While both subclonal and clonal mutations were significantly associated with the response (Fig. 2C), the first ones had a larger effect size. The odds ratio obtained from a logistic regression was higher for subclonal TMB (OR = 1.55, 95% CI [1.30; 1.92]) than for clonal TMB (OR = 1.24; 95% CI [1.10; 1.42]). This is in contrast to previous findings based on one of the cohorts, HdM-BLCA-1<sup>18</sup>, in which clonal mutations were shown to have a larger influence on the response. We also observed a significant enrichment of APOBEC-induced mutations in the responder group (Fig. 2D); this variable was moderately correlated with TMB ( $r = 0.47$ , Fig. 2E). The two APOBEC signatures SBS2 and SBS13 were significantly associated with the response, but the first one had a stronger effect (linear regression values  $B = 5.40$  and  $p$  value = 0.0048) than the second one ( $B = 1.91$  and  $p$  value = 0.0107).

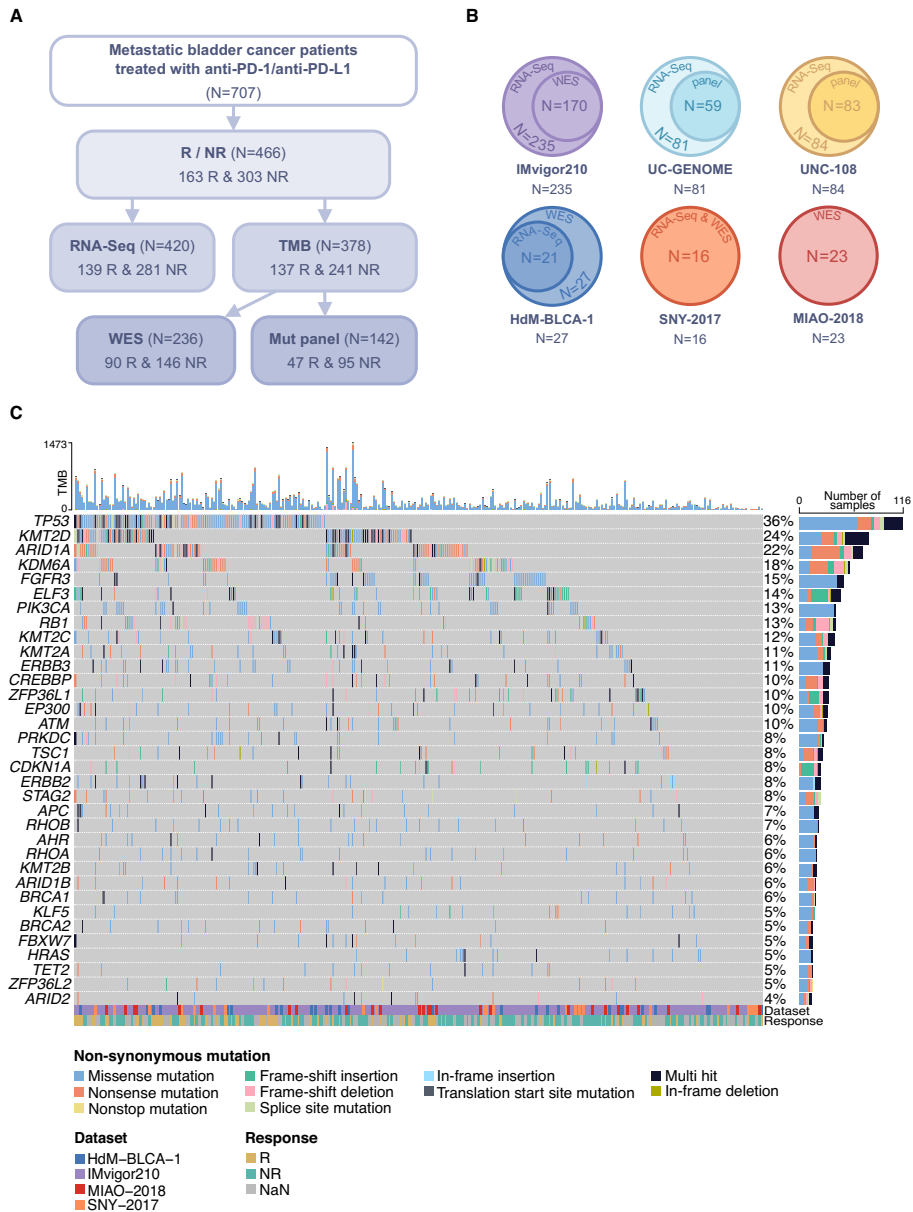
Missense mutations were significantly associated with response (Fig. 2F, Supplementary Table 3), independently of the predicted impact of the mutation on protein function (Supplementary Fig. 1). We found that non-stop mutations (single nucleotide variants that cause a stop codon loss) were significantly associated with response, but this was not the case for frameshift or in-frame insertions and deletions (INDELs). We also predicted neoantigens derived from missense mutations using the NetMHCpan software<sup>23</sup>. As previously described for different cancer types<sup>6,18,24</sup>, we found that the number of predicted neoantigens derived from missense mutations was significantly associated with response. Finally, we constructed peptides resulting from frameshift INDELs and the protein extensions generated by nonstop mutations. For these mutations we found no difference in the number of putative neoantigens between responders and non-responders (Supplementary Fig. 1).

Somatic copy number alterations are common events in tumor cells of BLCA patients<sup>10</sup>. Previous studies have suggested a positive relation between focal amplification of cyclin D1 (*CCND1*) and ICI response<sup>15</sup> and a negative association between genomic alterations in *CDKN2B* and *CDKN2A* and overall survival in BLCA patients<sup>25,26</sup>. While we found *CDKN2A/CDKN2B* to be the most common deletion in our data (16 responders and 30 non-responders), and every second patient to have an amplification in the *CCND1* gene region (51.30%, 75 non-responders and 43 responders), none of the copy number alterations were found to be significantly associated to ICI response (Chi-square test  $p$  value > 0.05).

### Gene expression biomarkers associated with ICI response

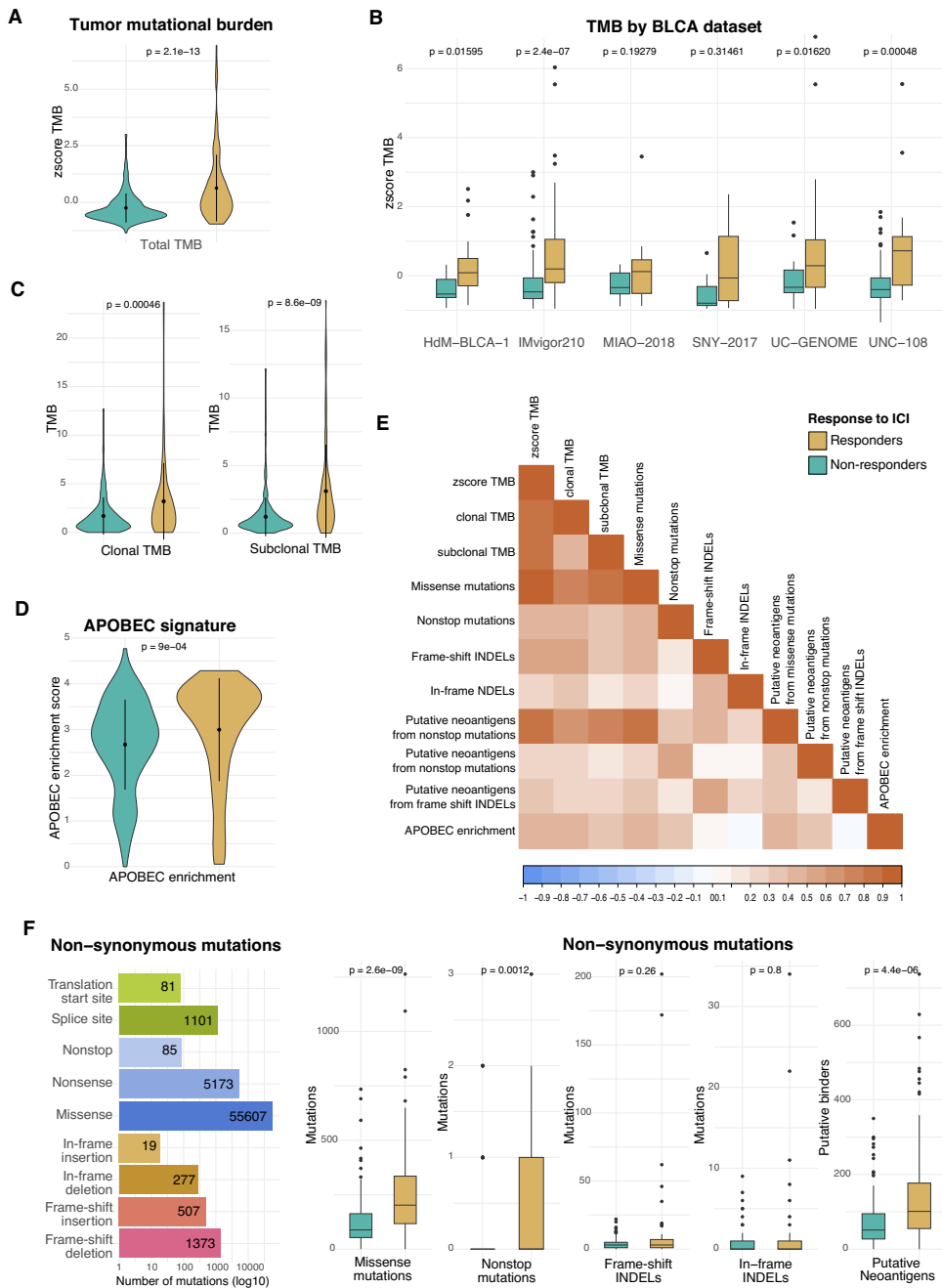
We used the RNA-Seq data to compute individual gene expression values, immune cell composition in the tumor sample with the CIBERSORTx algorithm<sup>27</sup>, and previously defined gene signatures. As expected, immune invasion and activation biomarkers were consistently more elevated in responders than in non-responders. This included chemokines such as CXCL9, CXCL10, CXCL11, and CXCL13, the interferon- $\gamma$  and granzyme A and B, and CD8A (Fig. 3A and Supplementary Fig. 1). We observed significantly higher expression of the immune checkpoints PD-1 and PD-L1 in responders than in non-responders (Fig. 3B). While high expression of these genes is generally associated with poorer prognosis in cancer, they have also been suggested as biomarkers of ICI response<sup>28–31</sup>.

The abundance of CD8+ T cells, memory-activated CD4+ T cells and pro-inflammatory macrophages M1 was significantly higher in responders than in non-responders (Fig. 3C). Three out of four gene expression signatures related to CD8+ T cell activation were significantly associated with response: CD8 T-effector from Mariathasan et al.<sup>9</sup> for bladder cancer, CD8 T-effector from McDermott et al.<sup>32</sup> for kidney cancer and CD8 T-effector from POPLAR for lung cancer<sup>33</sup> (Fig. 3D). CD8 T-effector from Bindea et al.<sup>34</sup> for colorectal cancer did



**Fig. 1 | Overview of the analyzed sequencing data of six metastatic bladder cancer datasets. A** Number of patients with different types of data. Sequencing data of 707 patients was downloaded. Patients with a RECIST classification of partial or complete response to anti-PD-L1/PD-1 were considered responders (R) and progressive disease were considered non-responders (NR). Stable disease and non-evaluable were not included for further analysis. **B** Of 466 patients with response to immunotherapy information, 348 patients have

RNA-Seq data and a TMB estimate available. **C** Most frequently mutated BLCA genes mutated in  $\geq 5\%$  of the samples over all four datasets with WES data (N = 318 patients). The X axis represents the patients from the different cohorts, the corresponding TMB is indicated. Mutations are classified in different types as indicated; multi hit describes cases where the gene carries more than one mutation in the patient. R responders, NR non-responders, Mut panel DNAmutation panel.



not achieve statistical significance. Similarly to the gene expression of IFNG, responder samples were significantly enriched in both signatures of the IFN- $\gamma$  pathway by Ayers et al.<sup>35</sup> and Cristescu et al.<sup>36</sup>. The recently published gene signature VIGex, which combines different genes involved in immune response<sup>37</sup>, and Neut-IFN-15, relating IFN- $\gamma$  stimulated neutrophil to improved response to ICI<sup>38</sup>, were also

significantly associated with response (Fig. 3D). We observed higher scores for the antigen-presenting machinery (APM) gene signature by Thompson et al.<sup>39</sup>, as well as for HLA-I, in responders than in non-responders (Fig. 3E and Supplementary Fig. 2).

We also investigated if tumor-restricted lncRNAs, which might contain non-canonical ORFs with the potential to generate HLA-bound

**Fig. 2 | Relationship between somatic mutations and the response to ICI.**

**A** Responders have a higher tumor mutational burden (Z score TMB) than non-responders (median Z score TMB R = 0.18, NR = -0.45). TMB is calculated as the number of non-synonymous mutations per 50 Mb. The differences were highly significant (two-sided, two-sample Wilcoxon test,  $p$  value =  $2.1 \times 10^{-13}$ ). **B** Responders show a higher mean TMB than non-responders. This finding is consistent over all six datasets, and four of the six datasets display significant differences (two-sided, two-sample Wilcoxon test). **C** The difference between treatment response groups remains when separating the TMB into clonal and subclonal by a cancer cell fraction (CCF) cutoff of 0.9 (two-sided, two-sample Wilcoxon test). Median clonal TMB R: 2.15 mut/50 Mb, NR: 1.1 mut/50 Mb; median subclonal TMB R: 1.19 mut/50 Mb, NR: 0.8 mut/50 Mb. **D** Responders are significantly enriched in APOBEC-induced mutations compared to non-responders (two-sided, two-sample Wilcoxon test,  $p$  value =  $9 \times 10^{-4}$ ; median APOBEC-enrichment score R: 3.45, NR: 2.85). **E** Spearman

correlation between different DNA-derived variables. **F** Number of non-synonymous mutations by type and association of the different mutation types with response. Missense mutations, nonstop mutations, and putative neoantigens were found to be significantly associated with the response (two-sided, two-sample Wilcoxon test). Medians for number of missense mutations R: 202 muts, NR: 88 muts; nonstop mutations R: 0 muts, NR: 0 muts, frameshift insertions/deletions R: 3 muts, NR: 3 muts; in-frame insertions/deletions R: 0 muts, NR: 0 muts; putative neoantigens R: 22.5 neoantigens, NR: 12 neoantigens. Neoantigens were predicted from missense mutations applying a threshold of 500 nM IC<sub>50</sub> binding affinity in NetMHCpan 4.0.  $N$  = 234 patients. R responders (yellow), NR non-responders (turquoise), TMB tumor mutational burden, INDELs insertions and deletions. All  $p$  values are indicated in the according plots. Source data are provided as a Source Data file.

peptides<sup>40,41</sup>, were more prevalent in responders than non-responders. If so, this would support the hypothesis that non-coding sequences might be a source of immunogenic antigens<sup>42</sup>. In each cohort, we generated a gene expression signature based on tumor-specific lncRNAs containing open reading frames with cancer-derived immunopeptidomics evidence (Supplementary Fig. 3). We found that, when the signatures from different datasets were combined together, responders had significantly higher values of this lncRNA signature than non-responders (Fig. 3F). When the analysis was performed separately for each cohort, only IMvigor210 achieved statistical significance (Supplementary Fig. 4).

We found significantly lower expression of the immunosuppressive transcription growth factor beta 1 (TGF- $\beta$ ), and of the cyclin gene CCND1, in responders than in non-responders (Fig. 3A). A biological pathway supporting tumor progression is the epithelial to mesenchymal transformation in stroma cells (Stroma/EMT)<sup>9,43</sup> was also negatively associated with response (Fig. 3D).

**Relationship between different signatures**

The different immune activation signatures tested were all strongly positively correlated with each other (Fig. 3G, Supplementary Data 1). Intriguingly, the immunosuppressive stroma/EMT signature, which was weakly positively correlated with immune activating signatures, showed a negative correlation with TMB ( $r$  = -0.169,  $p$  value = 0.002). CCND1, more abundant in non-responsive tumors, was positively correlated with TGF- $\beta$  ( $r$  = 0.177,  $p$  value < 0.001). PD-1 and PD-L1 expression levels, as well as HLA-I and HLA-II signatures, showed a positive correlation with CD8+ and CD4+ activating cells ( $r$  = 0.4–0.65,  $p$  value < 0.001). The lncRNA signature showed no correlation with the other biomarkers studied. Principal component analysis identified five major groups of variables that contribute to explain the variability across tumor samples: 1. Mutation status (TMB/APOBEC), 2. Immune activation signatures (CD8+ T cells, IFN- $\gamma$ , etc.), 3. TGF- $\beta$ /EMT immuno-suppressor signatures; 4. tumor-specific lncRNAs and, 5. CCND1 (Supplementary Fig. 5).

**The determinants of the response to ICI response depend on the subtype**

The analysis of BLCA samples from the TCGA project identified five subtypes based on different biomarkers<sup>10</sup>. We used this well-established methodology to classify the tumors in our meta-cohort. The largest group of patients was luminal-infiltrated (52R & 116 NR), followed by basal-squamous (35R & 78 NR), luminal-papillary (31R & 56 NR) and luminal (11R & 25 NR) (Fig. 4A). Neuronal represents the smallest subtype (10R & 6 NR). The only subtype with a significantly better response to treatment than the general trend was neuronal (62.5% versus 33% of responders, Fisher exact test,  $p$  value = 0.014).

We observed that 30% of luminal-papillary tumors were *FGFR3*-altered (26 out of 87), followed by 8% for luminal. Basal-squamous and luminal-infiltrated had high immune cell infiltration, as expected<sup>10</sup>

(Figs. 4B and 4C, respectively; Supplementary Fig. 6). In addition, we observed that these two subtypes had higher HLA-I gene expression (Fig. 4D) and, at least for basal-squamous, higher expression levels of the antigen presentation pathway (Fig. 4G). They also had high TGF- $\beta$  and low TMB values (Fig. 4D–F).

We next analyzed the different variables in the immune-infiltrated (luminal-infiltrated and basal-squamous) and non-immune-infiltrated (luminal-papillary, luminal and neuronal) subtypes separately. Although immune-infiltrated tumors had significantly lower TMB than non-immune infiltrated ones (Fig. 5A), TMB and APOBEC enrichment were significantly associated with response (Fig. 5B). The effect of immune activation and immunosuppression signatures on the response was significant in both groups, but in general more marked in the case of the immune-infiltrated subtypes (Fig. 5C, D, Supplementary Figs. 7 and 8). For the IFN- $\gamma$  signature, the difference between the median of the response group and the median of the non-response group was 0.727 for immune-infiltrated and 0.229 for non-immune-infiltrated. We observed the same trend for the CIBERSORT score (0.364 and 0.112, respectively), M1 macrophages (0.099 and 0.037, respectively) and CD8 T cells (0.07 and 0.026, respectively). For lncRNAs, the signature was only significant for the immune-infiltrated group (Fig. 5E).

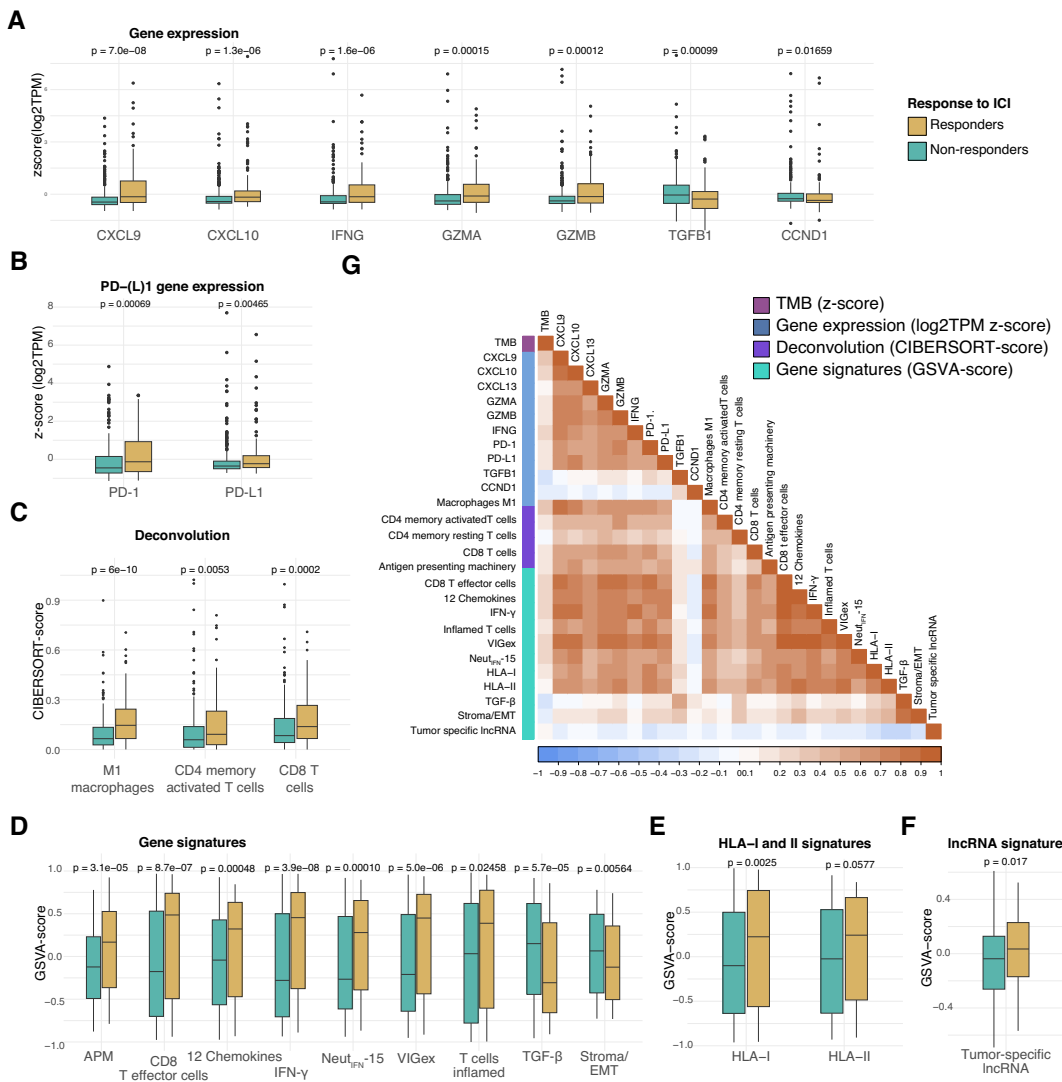
Because of the high TGF- $\beta$  expression values detected in the immune-infiltrated types (Fig. 5C), we decided to investigate in more detail the relationship between immune activation and immune suppression signatures. We found a relatively large group of patients with high CD8+ T cell infiltration and also high TGF- $\beta$  values, these patients had a lower probability to respond than those with high CD8+ and low TGF- $\beta$  values (39.7% versus 52.4% of responders, Fisher exact test  $p$  value = 0.0479, Fig. 5F). When we performed the analysis per subtype, the same trend was observed, especially for luminal-infiltrated, although the results did not achieve statistical significance (Supplementary Fig. 9). These observations can explain why tumors with high immune infiltration respond less well to the treatment than expected.

**Predictive models integrating different omics variables**

Based on the above-described findings, we selected a list of representative variables to construct predictive models of the response to ICI. The complete set of variables comprised mutation-based variables (TMB, non-stop mutations, APOBEC-enrichment score), gene expression of selected genes (*PD-1*, *PD-L1*, *CCND1*), immune signatures (IFN- $\gamma$ , stroma/EMT, inflamed T cells, TGF- $\beta$  and antigen-presenting machinery pathways), a signature of tumor-specific lncRNAs, immune cell abundance (M1 macrophages, CD4 memory activated T cells, CD8 T cells and regulatory T cells), as well as clinical information (ECOG, liver metastasis).

A random forest model trained with these variables achieved high accuracy, as indicated by an area under the curve (AUC) of 0.761 (Fig. 6A, Supplementary Table 4). In clinical practice, the number of mutations in a tumor sample is a widely used marker for decision-making in the context





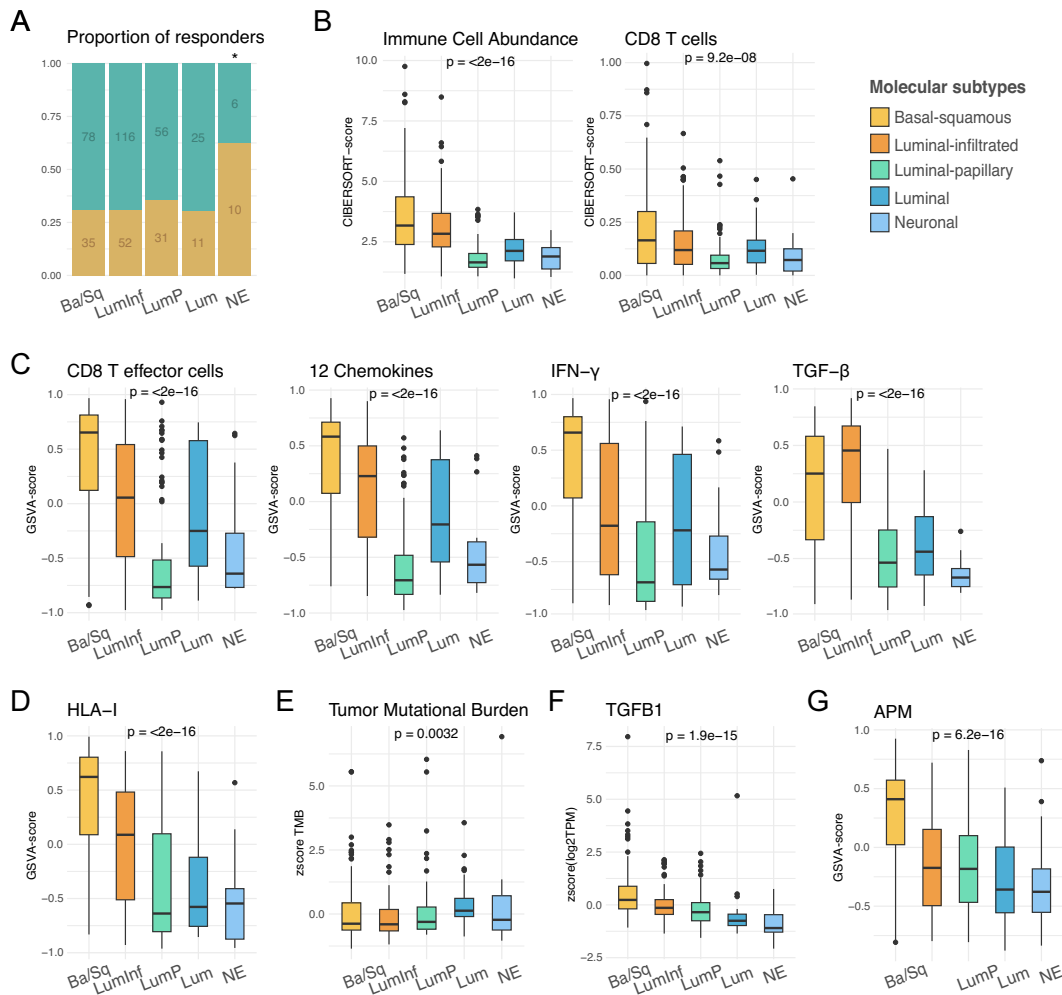
**Fig. 3 | Dissecting the effects of different immune-related variables.**

**A** Normalized expression values for selected genes described in the context of Immune activity or suppression. All comparisons were significant after multiple testing adjustment, except for the genes *TGFB1* and *CCND1* (two-sided, two-sample Wilcoxon test). **B** Expression of the immune checkpoint molecules PD-L1 (gene *CD274*) and PD-1 (gene *PDCDI*) is significantly higher in responders (two-sided, two-sample Wilcoxon test). **C** Deconvolution analysis using CIBERSORT shows higher immune cell abundance in responders (two-sided, two-sample Wilcoxon test). **D** Six signatures of tumor antigen presentation and immune response are enriched among responders, while immune suppression signature scores are higher in non-responders. All comparisons were significant after multiple testing adjustment,

except T cells inflamed and Stroma/EMT (two-sided, two-sample Wilcoxon test). **E** Gene signatures combining the expression of HLA-I and HLA-II types are higher in responders than in non-responders (two-sided, two-sample Wilcoxon test). **F** Responders show higher expression of a signature related to tumor-specific long non-coding RNAs (lncRNA) than non-responders (two-sided, two-sample Wilcoxon test). **G** Correlation matrix including TMB and RNA-Seq variables. *P* values obtained by two-sided, two-sample Wilcoxon test. *N* = 420 patients. R responders (yellow), NR non-responders (turquoise), APM antigen-presenting machinery, lncRNA long non-coding RNA. All *p* values are indicated in the according plots. Source data are provided as a Source Data file.

of ICI therapy<sup>44</sup>. A random forest model considering only Z score TMB had an AUC of 0.678. A threshold model in which tumor samples with more than 10 mutations/Mb were predicted to be from responders and those with less mutations from non-responders was associated with an AUC of 0.61 (Supplementary Fig. 10).

Random forest models provide information on the contribution of each variable to the model (feature importance). The variables with a clear association with response in the complete model were, in descending order, TMB, M1 macrophages, APOBEC-enrichment, IFN- $\gamma$  signature, CD8+ T cell abundance, PD1 gene expression and HLA-I



**Fig. 4 | The TCGA-subtypes luminal and neuronal have higher TMB, while basal-squamous and luminal-infiltrated show higher immune activity.** **A** Proportion of responders (yellow) and non-responders (turquoise) in each of the five TCGA subtypes. TMB was weakly correlated with the immune activation signatures. The neuronal subtype is the only one with a significant excess of responders over non-responders (two-sided Fisher's exact test,  $p$  value = 0.014, indicated with \*). **B** Basal-squamous and luminal-infiltrated have high values of immune cell abundance (absolute score obtained from CIBERSORT), while luminal, luminal-papillary and neuronal show low immune invasion. Immune cell abundance of CD4 T cells is similarly high in basal-squamous and luminal-infiltrated subtypes (CIBERSORT). Luminal-infiltrated shows the highest mean of CD4 memory resting T cells (CIBERSORT) (Kruskal–Wallis test). **C** Basal-squamous and luminal-infiltrated have

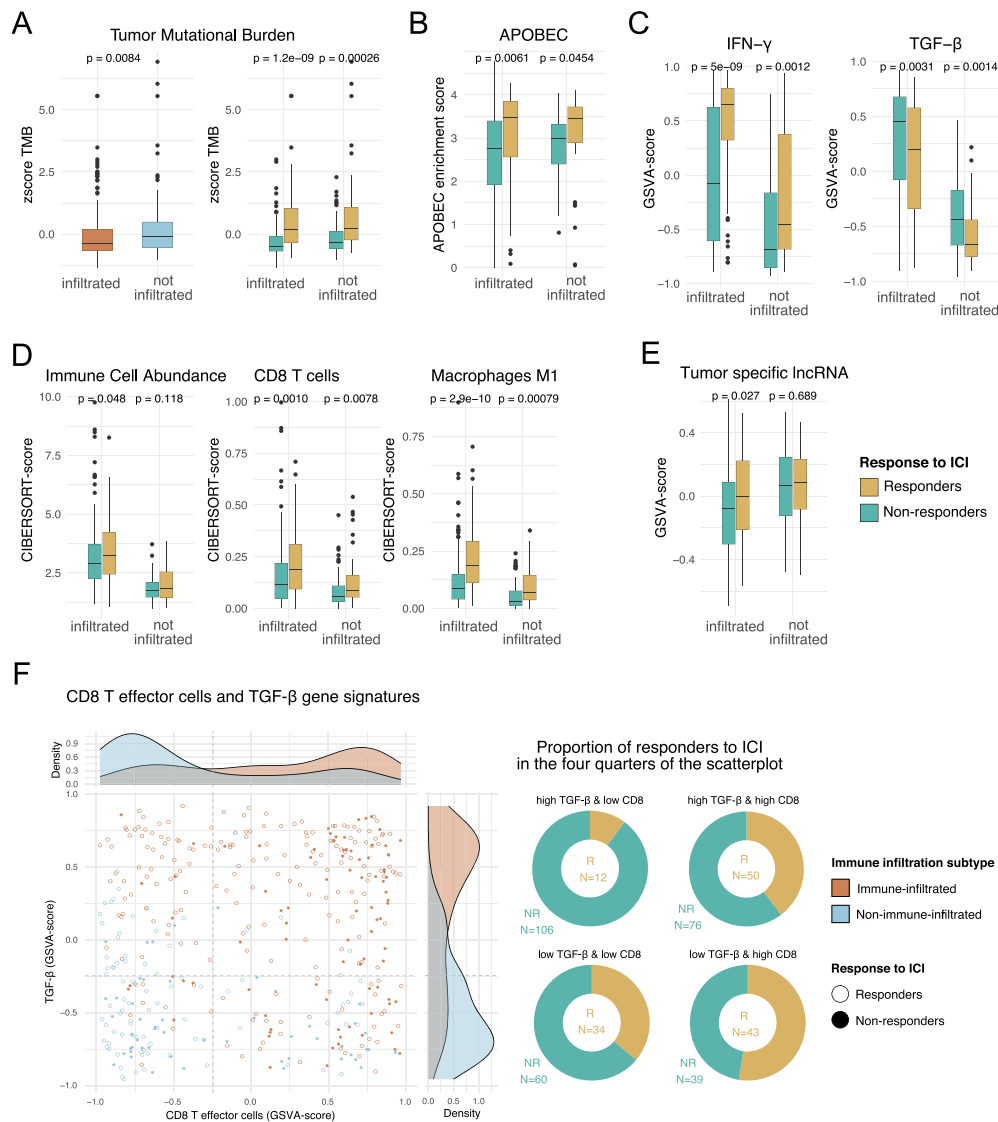
the highest expression of immune biomarkers, both immune activating and suppressive markers (Kruskal–Wallis test). **D**. HLA expression is also highest in luminal-infiltrated and basal-squamous (Kruskal–Wallis test). **E** TMB is highest in the luminal and neuronal subtypes (Kruskal–Wallis test). **F** The tumor-infiltrated subtypes basal-squamous and luminal-infiltrated show the highest values of TGF- $\beta$  gene expression (Kruskal–Wallis test). **G** Basal-squamous is enriched in antigen-presenting machinery (APM) compared to other subtypes (Kruskal–Wallis test).  $N = 420$  patients. Ba/Sq: Basal-squamous (yellow), LumInf Luminal-infiltrated (orange), LumP Luminal-papillary (green), Lum Luminal (dark blue), NE Neuronal (light blue). All  $p$  values are indicated in the according plots. Source data are provided as a Source Data file.

signature (feature importance > 0.03, Fig. 6B). Those associated with lack of response were, also in descending order, *CCND1* gene expression, Stroma/EMT signature and TGF- $\beta$  gene signature.

As a validation cohort, we tested the model on data from the JAVELIN Bladder 100 trial<sup>45</sup>, which includes advanced urothelial cancer patients treated with avelumab (anti-PD-L1). While no raw data was available, we were provided with processed data. With this data, we tested the model based TMB + RNA using data from 123 patients from

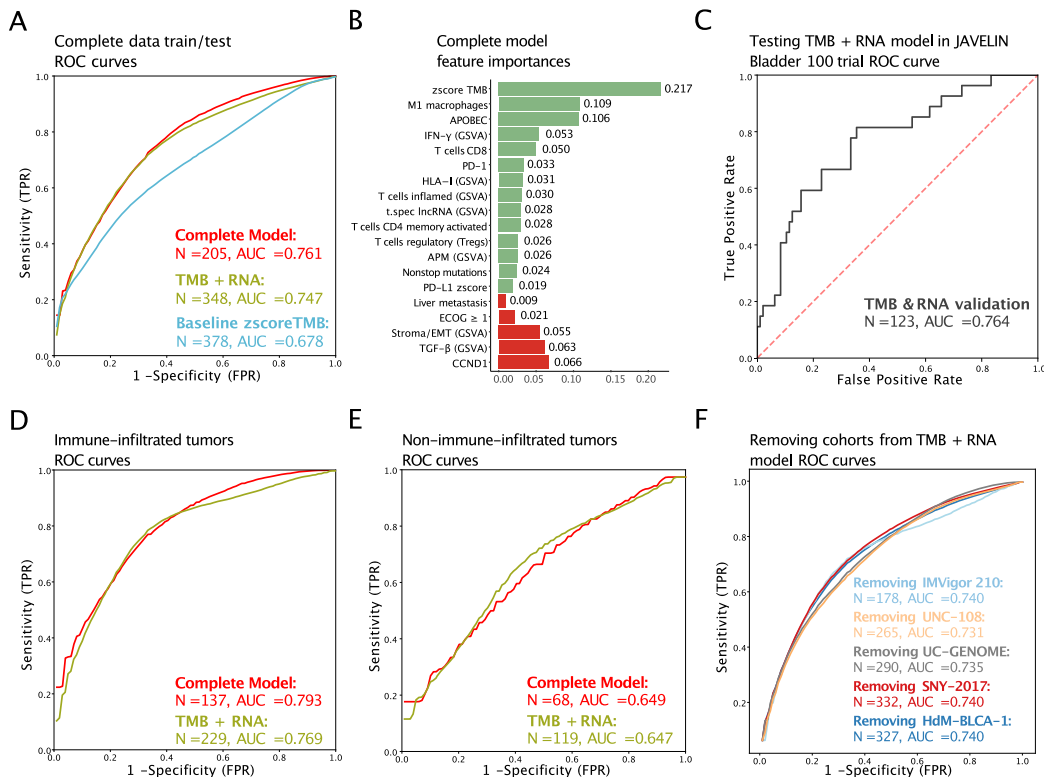
the JAVELIN Bladder 100 trial (27 responders, 96 non-responders). The model achieved an AUC of 0.764 in the validation run, surpassing the averaged AUC of 0.747 obtained from 1000 seeds in the train/test phase (Fig. 6C). This result underscores the robustness of our model, highlighting its potential in predicting outcomes for patients treated with immune checkpoint inhibitors.

We then built models separately for the immune-infiltrated and the non-immune-infiltrated classes. Maximum accuracy was achieved



**Fig. 5 | Influence of different biomarkers in the immune-infiltrated and non-immune-infiltrated subtypes. A** Overall, TMB is higher in the non-immune infiltrated group. In both groups, responders have significantly higher TMB (two-sided, two-sample Wilcoxon test). **B** In both groups, responders have higher enrichment scores of the APOBEC mutational signature (two-sided, two-sample Wilcoxon test). **C** Responders have higher IFN- $\gamma$  expression values, and lower TGF- $\beta$  expression values than non-responders for both subgroups (two-sided, two-sample Wilcoxon test). **D** CIBERSORT score is higher in responders of the infiltrated groups, but not the non-infiltrated group. Differences between responders and non-responders are significant for CD8+ T cells and macrophages M1 (two-sided, two-sample Wilcoxon

test). **E** Only in the infiltrated group, responders have higher expression of the lncRNA signature. No difference was observed between response groups in the non-immune-infiltrative group (two-sided, two-sample Wilcoxon test). **F** Relationship between CD8 T cell and TGF- $\beta$  gene expression. Immune-infiltrated samples tend to have high CD8+ T cell abundance, and in many cases also high levels of the TGF- $\beta$  signature. The dashed line marks the optimal cutoff for each gene signature obtained by ROC (TGF- $\beta$ : 0.0163; CD8 t effector cells: 0.246), responder are marked with circles, non-responder with filled circles.  $N = 420$  patients. R: responders, NR: non-responders. All  $p$  values are indicated in the according plots. Source data are provided as a Source Data file.



**Fig. 6 | Predictive models of the response to immunotherapy.** **A** ROC curves for the complete model, TMB and TMB + RNA (TMB and RNA-seq derived variables), with AUC and number of samples (*N*). FPR: false positive rate; TPR: true positive rate. The average of 1000 runs is shown. **B** Feature positively or negatively associated with response. The length of the bar represents feature importance from random forest. Color reflects association with response taken from the previous manuscript sections (green: positively associated, red: negatively associated). **C** ROC curve for testing the TMB + RNA model in the validation cohort JAVELIN Bladder 100 trial. AUC and *N* values are indicated. **D** ROC curve of the complete model and TMB + RNA for the subset immune-infiltrated tumors, with AUC and *N* values. **E** ROC curve of the complete model and TMB + RNA for the subset of non-immune-infiltrated tumors, with AUC and *N* values. **F** ROC curves for TMB + RNA model when removing one of the cohorts at a time.

in the immune-infiltrated group using all the variables (Fig. 6D, AUC = 0.793). The IFN- $\gamma$  signature, as well as *CD8* gene expression, gained weight in this model with respect to the model developed for all patients (Supplementary Fig. 11). *CCND1* abundance was the most negative factor, followed by TGF- $\beta$  levels. In contrast, the non-immune-infiltrated model showed relatively low accuracy (Fig. 6E, AUC = 0.649). The model based on RNA-Seq data plus TMB had an AUC of 0.769 for immune-infiltrated (*N* = 229) and 0.647 for non-immune-infiltrated (*N* = 119), similar to the complete model.

The robustness of the model was further tested by removing one dataset at a time. For this, we employed the model constructed with TMB and RNA-Seq data, as it included more patients than the complete model (348 *versus* 205) and showed almost the same accuracy (AUC 0.747 *versus* AUC 0.761, Fig. 6A). Removing individual datasets resulted in models of similar accuracy (Fig. 6F), indicating that the model does not suffer from over-fitting. The fact that the number of non-responders was approximately double than the number of responders did not seem to have a significant effect either (Supplementary Fig. 12).

To gain further insights into the drivers of the response to ICI in the non-immune-infiltrated samples, we derived subtype-specific decision trees and inspected the distribution of the variables across patients (Supplementary Figs. 13 and 14). In the case of luminal-papillary, the largest group, we also built a random forest model (*N* = 71; Supplementary Table 5). This model identified PD-1 Z score as the third most important feature following macrophages M1 and TMB. The decision tree also indicated that high PD-1 expression was strongly associated with response, with 11 out of 14 patients with PD-1 Z score > -0.41 being responders. Among the remaining 15 responders, 13 had very low values of the stroma/EMT gene signature (stroma/EMT < -0.527). Therefore, a large part of the responders, 24 out of 26, were characterized by either high PD-1 values or low stroma/EMT.

In the case of luminal, the decision tree that we obtained revealed an association between absence of Regulatory T cells (Treg) and response to ICI. The percentage of responders with Treg values above 0.003 was only 9% (1 out of 11), whereas it was 58% for non-responders (14 out of 24) (*p* value = 0.0095). Regulatory T cells are a

subset of CD4<sup>+</sup> T cells with immunosuppressive effects<sup>46</sup>. This effect was only observed in the luminal subtype (Supplementary Fig. 14).

In the neuronal subtype, the lack of response was invariably associated with low AMP signature values (AMP Z score < −0.36 in all non-responders, see also Supplementary Fig. 8). So, the inability to present antigens at sufficiently high levels seems to be a strong determinant for the treatment to fail in this group of patients.

## Discussion

We have compiled the largest clinical, mutation and gene expression dataset from patients with advanced BLCA treated with ICI. Reanalysis of previously associated signatures has confirmed that tumor mutational burden (TMB) and APOBEC-induced mutations are consistently associated with the response<sup>3,15,18,47,48</sup>. Frameshifts mutations, reported to be significant in melanoma<sup>49</sup>, were not found to be significant here. Analysis of RNA sequencing data has confirmed that immune activation biomarkers were significantly associated with response to ICI, as previously reported for different cancer types<sup>15,35,36</sup>. Macrophage M1 and CD8<sup>+</sup> T cell levels were higher in responders than in non-responders, even in those cancer subtypes with relatively little immune infiltration. Two factors that negatively affected the response were TGF- $\beta$ /stroma/EMT and *CCND1* gene expression levels. It has been proposed that TGF- $\beta$  negatively affects the response to ICI by blocking T cell penetration in the tumor<sup>9</sup>. In the case of *CCND1*, a previous association had been found between *CCND1* amplification and lack of response<sup>45</sup>. The inhibition of the protein complex cyclin D1/CDK4 results in higher PD-L1 levels, linking the kinase activity modulated by *CCND1* with cancer immune surveillance<sup>50</sup>. Our results indicated that, at least in bladder cancer, measuring *CCND1* gene expression could have a stronger prognostic value than measuring copy number.

We found several unexplored biomarkers of the response to ICI. The first one was the number of non-stop mutations, which was positively associated with response. Although non-stop mutations are only present in about 21% of the tumors, the large number of patients analyzed here allowed us to draw this association. Non-stop mutations can extend the coding sequence and, as a result, “foreign” antigens might be formed. Another signature was linked to the tumor-specific activation of lncRNAs, which are an alternative source of tumor-specific antigens<sup>42,51</sup>. Mutations in the gene *ARHGEF12* were also significantly associated with response. This gene encodes a yet poorly characterized Rho GTPase. Interestingly, a study comparing pre- and posttreatment samples of two cancer patients treated with anti-PD-1 reported that tumor clones with *ARHGEF12* mutations were only detected in pretreatment samples<sup>52</sup>. This could imply that anti-PD-1 treatment was effective in eliminating cells containing mutated *ARHGEF12*.

Previous models of the response to immunotherapy included data from different types of cancer<sup>15,17,53,54</sup> or focused on specific BLCA cohorts<sup>16</sup>. The size of our study allowed us to investigate the determinants of the response to ICI in different BLCA subtypes. Surprisingly, we found that the two subtypes with the highest immune infiltration - basal-squamous and luminal-infiltrated - did not show an overall better response than the other subtypes. This is likely to be relevant for trials that select treatment on the basis of cancer subtype<sup>55</sup>. We discovered that the immune-infiltrated subtypes tend to have lower TMB values, as well as higher immunosuppressive signatures, than other subtypes, and these factors decrease the likelihood of responding. For immune-infiltrated subtypes we achieved high accuracy in response prediction models that included RNA-Seq data in addition to TMB. This encourages the use of expression biomarkers, such as signatures of macrophages M1, CD8<sup>+</sup> T cells and TGF- $\beta$ , in predicting the outcome of the treatment for these cases. For non-immune-infiltrated subtypes, other biomarkers appeared to be particularly relevant. PD-1 expression levels were strongly associated with response in the case of luminal-papillary and basal-squamous, but were of little relevance in the other

subtypes. The potential inhibitory effect of Treg cells was consistent with the larger number of these cells in the tumors of non-responders versus non-responders, in the case of the luminary subtype.

Anti-PD-(L)1 therapy depends on the expression of PD-1 and PD-L1, which can be quantified using immunocytochemistry prior to treatment<sup>2,56</sup>. The expression of these genes was significantly associated with response. However, it was also more elevated in the immune-infiltrated tumors than in the non-immune-infiltrated tumors, independently of the response. This illustrates how taking into account the specific tumor subtype, or the degree of immune inflammation of the tumor, might be key for a correct interpretation of these biomarkers in a clinical setting.

Although a large number of variables were considered here, there may be yet unknown factors that also have an influence in the response. The neuronal subtype showed a strikingly high response rate. It has been hypothesized that this kind of tumor could be more responsive to ICI due to the expression of tissue-restricted neuronal or neuro-endocrine proteins, together with low TGF- $\beta$  gene expression values<sup>44</sup>. The present study, which included a larger number of patients, confirmed that TGF- $\beta$  levels are particularly low in this class. As already mentioned, the non-responders (6 out of 16) were all characterized by low values of the antigen-presentation machinery. This reinforces the fundamental role of tumor-specific antigens, of mutation origin or otherwise, in mediating the response to the treatment.

Using part of the data for validation we obtained an AUC estimate of 0.761 for the model that used the complete set of variables (0.793 for the immune-infiltrated subtypes) and 0.747 for the RNA + TMB model (gene expression variables and TMB). This represented an improvement over using TMB alone (0.678). When we tested the RNA + TMB model against the JAVELIN Bladder 100 trial we obtained an AUC of 0.764. These predictive values are in the range of those obtained by a previously obtained model that combined data from different types of cancer<sup>15</sup>; the AUC values were between 0.66 and 0.86 depending on the test dataset. The elastic net logistic regression model by Damrauer et al.<sup>16</sup>, built on the bladder cancer IMvigor210 dataset, showed an AUC of 0.84 when tested on the validation portion of the same dataset, 0.82 when tested on UNC-108 and 0.65 when tested on UC-GENOME.

In this study we integrated disparate cohorts of bladder cancer patients, and this poses a number of challenges. In order to control for sequencing batch effects we transformed the data to normalized Z scores, which allowed us to merge datasets. Not all the datasets included the same set of variables, therefore, the complete model did not include all patients with ICI response information. We tested the effect of leaving out one cohort of a time, or employing a similar number of non-responders than responders. External cohort validation using the independent JAVELIN Bladder 100 cohort demonstrated the robustness of our model. In this specific cohort the patients had been treated with chemotherapy prior to immunotherapy, but this does not seem to impact the results. Testing the model in cohorts treated with anti-PD-(L)1 in combination with other drugs (e.g. Enfortumab vedotin) can be foreseen.

In conclusion, we have combined six datasets to build the largest bladder cancer-specific meta-cohort of ICI therapy, and conducted a comprehensive multi-omics analysis. The results have uncovered unexplored ICI associated variables and have shed light on the complex interplay between tumor biology, the immune microenvironment and treatment response. Our study underlines the importance of subtype-specific factors for more personalized treatment strategies and enhanced patient outcomes in the era of immunotherapy.

## Methods

### Omics datasets

We accessed tumor RNA-Seq, paired tumor-germline WES and targeted DNA sequencing from six advanced or metastatic bladder cancer

cohorts. Ethical approval to use the data for research was obtained by the original institutions who published the original data used in this study. All patients received anti-PD-1/PD-L1 treatment. Germline WES was conducted on whole blood samples, tumor WES and RNA-Seq data obtained from FFPE samples coming from primary or metastatic cancer tissue.

1. IMvigor210,  $N = 348$ , Mariathasan et al.<sup>9</sup>

EGA study ID EGAS00001002556 [<https://ega-archive.org/studies/EGAS00001004343>]

2. HdM-BLCA-1,  $N = 27$ , Boll et al.<sup>18</sup>

EGA study ID EGAS00001007086

3. SNY-2017,  $N = 25$ , Snyder et al.<sup>3</sup>

dbGaP accession ID phs001743 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001743.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001743.v1.p1)]

4. MIAO-2018,  $N = 26$ , Miao et al.<sup>2</sup>

dbGaP accession ID phs001565 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001565.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001565.v1.p1)] and phs000694 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000694.v3.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000694.v3.p2)]

5. UC-GENOME,  $N = 191$ , Rose et al.<sup>19</sup>

dbGaP accession ID phs003066 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs003066.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003066.v1.p1)]

6. UNC-108,  $N = 89$ , Damrauer et al.<sup>16</sup>

Omnibus accession ID GSE176307

Three patient samples were excluded due to corrupted raw data files (IMvigor210: 10059, 10119, SNY-2017: 9881). Patient 4072 of the SNY-2017 cohort was excluded due to low tumor mean coverage according to the publication.

As a validation cohort, data of the JAVELIN Bladder 100 cohort (ClinicalTrials.gov identifier: NCT02603432)<sup>15</sup> was obtained from Merck Healthcare KGaA under a data sharing agreement. The dataset is derived from biological samples of advanced bladder cancer patients who received chemotherapy, randomly assigned to best supportive care alone or in combination with the anti-PD-1 drug avelumab. From this cohort, we included 123 patients ranked as responder or non-responder to immunotherapy following RECIST1.1 criteria.

### Single nucleotide variant calling

All paired tumor and germline WES samples were processed using the same pipeline following the GATK pipeline (version 4.0.1.2). Raw sequencing reads were trimmed using cutadapt (version 4.1), the quality was controlled using fastQC (version 0.11.7). Reads were aligned to the reference genome GRCh38 using BWA (version 0.7.17), and, in cases of several fastq files per patient, merged into one BAM file using GATK MergeSamFiles. Next, intervals and INDELS were realigned using the GATK RealignerTargetCreator and IndelRealigner to finally calibrate base scores and mark duplicates using GATK BaseRecalibrator. A summary file of the final BAM file was built using GATK GetPileupSummaries, contamination was estimated with GATK CalculateContamination and coverage using qualimap (version 2.2.1). A metrics file was built using GATK CollectAlignmentSummaryMetrics. Mutations were called with the three callers GATK Mutect2, Strelka2 (version 2.9.10) and VarScan2 (version 2.4.4). We used SAMtools (version 1.12) mpileup to build the input for VarScan2. Additionally to the default filters, Mutect2 was run using the germline-resource file somatic-hg38.af-only-gnomad.hg38.vcf.gz from gatk hg38 and GATK FilterMutectCalls using the contamination file previously generated. For all callers only mutations flagged as PASS were kept. We then build an ensemble file of mutations detected by a minimum of two out of three callers using bcbio-variation ensemble (version 0.2.2.6) and annotate them with VEP (version 104). Maf files were generated using vcf2maf (version 0.1.16). Final filters applied to mutations were a population-wide allele frequency under 5% (gnomAD), a minimum sample depth of 30, and a minimum alternative allele depth of 3. Maftools R-package (version 2.10.0) was used for TMB calculation

(non-synonymous mutations per capture size of 50 Mb)<sup>57</sup>. For two patients of the IMvigor210 cohort, the variant callers detected zero somatic mutations (10124, 10299). In patient 6428 of the SNY-2017 dataset, none of the called mutations passed the filter. We identified a total of 64,275 somatic non-synonymous mutations in 15,439 genes. This represents an average of 202 non-synonymous mutations per patient (318 patients with WES data). If we only considered the 236 patients classified as responders or non-responders and with WES data, the average was similar (210 mutations per patient).

The selection of tested tumor suppressor and oncogenes was obtained from the following sources: COSMIC (<https://cancer.sanger.ac.uk/cosmic/census?tier=2>, last accessed September 11th 2023) as well as reported chromatin modifying genes, chromatin regulating genes and common bladder cancer genes. A complete list is provided in Supplementary Data 1.

### Somatic copy number alterations

The ASCAT R-package (version 3) was used to obtain tumor purity and local allele-specific copy number estimates from WES samples. ASCAT was run following the recommendations of version3 for WES data. AlleleCount (version 4.3.0) was used to generate LogR and BAF files. The necessary loci file was downloaded from the nf-core/sarek pipeline (release 3.1.) and subsetted for loci of the according BED file. We have tested differences between responders and non-responders using Chi-square tests for somatic copy number alterations of a selection of genes significantly amplified or deleted in TCGA (taken from Supplementary Table 2 of Zack et al.)<sup>58</sup>.

### Mutation clonality

Mutations with a cancer cell fraction (CCF) above 0.9 were considered clonal. CCF was calculated as:  $CCF = \frac{vaf}{p^*(2*(1-p) + c*p)}$ , with  $c$  being the copy number at the mutation position and  $p$  the tumor sample purity<sup>18</sup>. Patients with a tumor sample purity of 1 were excluded from the clonality analysis.

### APOBEC-enrichment estimation

To evaluate the enrichment of the apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) mutational signature, we used the trinucleotideMatrix function from the maftools R-package (version 2.10.0) assessing the enrichment of C>T mutations within TCW motifs relative to the overall count of C>T mutations compared against a background of cytosines and TCW motifs. Samples with an enrichment score exceeding 2 and a  $p$  value below 0.05 were deemed statistically significant. Additional mutational signatures were obtained with the deconstructSig R-package (version 1.9.0)<sup>59</sup> with COSMIC signatures and normalized by exome trinucleotide context (tri.counts.method = 'exome').

### Neoantigen and HLA prediction

The 4-digit HLA genotype corresponding to each patient was determined using the nf-core/hlatyping pipeline (release 1.2.0)<sup>60</sup> and the OptiType HLA genotyping algorithm (version 1.3.5)<sup>61</sup>. OptiType was executed with default parameters on blood samples to ascertain the HLA alleles of each patient. Subsequently, we generated all possible 9-mer peptides encompassing a mutation and their non-mutated counterparts. The peptide sequences were obtained from Ensembl GRCh38. The binding affinity of tumor peptides to MHC-I molecules was predicted using NetMHCpan (version 4.0)<sup>62</sup> applying an IC<sub>50</sub> threshold of 500 nM.

### Gene expression quantification

Total stranded RNA-Seq reads for all cohorts but UNC-108 were quality assessed using both FastQC (v0.11.5) and FastQScreen (v0.14.0) software 65. Bulk RNA-Seq sequencing reads from tumor samples were aligned to the human reference genome GRCh38 reference and the



gencode annotation (version 41) using STAR (version 2.7.8)<sup>63</sup>. We checked the strandedness of the dataset with the RSeQC program<sup>64</sup> and, according to the result, we used the appropriate mode of featureCounts<sup>65</sup>, from the Subread package (version 2.0.3), to quantify the gene expression. Normalization was performed by converting counts to transcripts per million (TPM). We integrated the information by using Z score transformation of log2(TPM) values in each of the datasets.

### Analysis of RNA-Seq data

Deconvolution values were obtained per dataset using the CIBERSORTx method on the matrices of TPM expression with the LM22 gene signature and B-mode batch correction in absolute mode<sup>27</sup>. A selected list of gene signatures previously described to be associated with ICI response was tested using the Bioconductor package GSVA<sup>66</sup>. GSVA scores were calculated per dataset and then merged for analysis. We computed the following gene signatures: CD8 T effector cell signature<sup>9</sup>, CD8 T effector cells from McDermott<sup>32</sup>, CD8 T effector cells POPLAR<sup>33</sup>, 12 chemokines<sup>67</sup>, IFN- $\gamma$ <sup>35</sup>, IFN-gamma from Cristescu<sup>36</sup>, Stroma and EMT<sup>43</sup>, inflamed T cells<sup>35</sup>, TGF- $\beta$ <sup>9</sup>, VIGex<sup>37</sup>, APM<sup>39</sup>, NeutIFN-15 (Interferon-stimulated neutrophils)<sup>38</sup> and signatures for HLA-I and HLA-II including the protein-coding HLA genes. A complete list of the genes included in the signatures can be found in the Supplementary Data 1.

The gene expression data was also used to classify the samples by molecular bladder cancer subtype BLCAsubtyping R-packages (version 2.1.1)<sup>10</sup>.

### Data normalization

We observed that the tumor samples from IMvigor210 had higher WES sequencing coverage than the other datasets while, among germline samples, HdM-BLCA-1 samples showed the highest coverage (Supplementary Fig. 15). These differences, however, did not influence the calling of mutations, as WES coverage and TMB were not correlated (Supplementary Fig. 15).

The purity of the tumor samples was quite homogenous across datasets, with a mean value of 0.569 (Supplementary Fig. 15). As expected, we identified a significant negative correlation between the tumor sample purity estimate obtained from ASCAT and the absolute immune cell infiltration score from CIBERSORT ( $r = -0.469$ ,  $p$  value  $< 0.001$ ).

The TMB estimates from the targeted DNA sequencing datasets UC-GENOME and UNC-108 had higher values than the estimates from WES data (Supplementary Figs. 16 and 17). To correct for this effect, we normalized the TMB values by dataset using Z scores.

### RECIST and clinical data

For the analysis of the association of biomarkers to immunotherapy response, we restricted the cohort to patients responding to the treatment (partial responders and complete responders) and not responding (progressive disease) following the RECIST1.1 criteria. Patients with a RECIST status of stable disease were not considered for the comparison of response groups. Although several studies have included the stable disease status as non-responders<sup>9,15,16</sup>, we aimed for a more conservative approach with the intention of a clearer separation between the response groups, as done previously by other groups<sup>36</sup>. Out of the initial 707 patients, there are 56 complete, 107 partial responders and 303 patients with progressive disease status. Clinical data was collected from the corresponding publications and data repositories. The clinical variables found to be significantly different between responders and non-responders using a two-sided, two-sample Wilcoxon test were ECOG  $\geq 1$  and liver metastasis. For 202 patients, no information on liver metastasis status was available and there was no ECOG evaluation for 44 patients.

### lncRNA signature

We generated a signature of tumor-specific long non-coding RNAs (lncRNA) expression (Supplementary Fig. 3). The lncRNAs included transcripts annotated as intergenic long non-coding RNAs as well as processed pseudogenes. Only lncRNAs with an expression above the 75% threshold for each dataset were considered. Gene expression data from a collection of 54 tissues from The Genotype-Tissue Expression (GTEx) project was collected. lncRNAs with significant expression (median expression higher than 0.5 TPM) in any non-reproductive tissue were discarded, resulting in the generation of lists of potential tumor-specific lncRNAs per cohort.

Additionally, we selected lncRNAs with matches in the publicly available cancer-associated immunopeptidomics dataset published by Chong and co-workers<sup>68</sup>. This restricted our initial list of lncRNAs to those generating peptides that could specifically bind HLA receptors in melanoma. This was done per dataset independently; a GSVA score per patient was calculated as a gene signature of the selected tumor-specific lncRNAs. The list of lncRNAs selected for each of the cohorts with RNA-Seq data can be found in Supplementary Data 1.

### Response prediction models

The machine learning framework developed for this project consists of three different components (Supplementary Fig. 18). The first one comprehends all the preprocessing steps such as scaling and encoding. In the training phase, the best hyperparameters for the ML algorithm are chosen via 15-fold-cross-validation and repeating this process multiple times (100 or 1000) in order to ensure that the parameters do not fit a particular seed. Following training, the model undergoes evaluation, where two different strategies are implemented. The first one is the commonly used train/test method. In our case, we separated the data by using a 70/30 split, leaving 30% of the data unseen by the predictive model. Using the randomly selected 70% we perform a 15-fold cross-validation for training the model. We tested with the remaining 30% of the data and obtained all the metrics (including AUC).

For the model we selected a set of variables that were found to be significantly associated with the response and which showed relatively low correlation among themselves and the least possible overlap in cases of gene signatures. The complete model includes the following variables: mutation-based variables (TMB (Z score muts/Mb), nonstop mutations (muts), APOBEC-enrichment score), gene expression of selected genes (Z score (log2TPM) of PD-1, PD-L1, CCND1), immune signatures (GSVA scores of IFN- $\gamma$ , stroma/EMT, inflamed T cells, TGF- $\beta$  and antigen-presenting machinery pathways) and a signature of tumor-specific and -expressed lncRNA (GSVA score), immune cell abundance (CIBERSORT scores of M1 macrophages, CD4 memory activated T cells, CD8 T cells and regulatory T cells), as well as clinical information (ECOG  $\geq 1$  (Y/N), liver metastasis (Y/N)).

We tested three different machine learning algorithms on our data to build a robust prediction model. Logistic regression with L1 regularization type was used as a baseline model. Another model that showed better performances than logistic regression and also provided positive and negative coefficients is the stochastic gradient descent. Finally, random forest was used for the final model. This method is very robust to outliers by averaging the scores of all the trees that comprehend the forest. Note that the feature importance that can be extracted from the random forest is based on the gini impurity decrease; hence, they are always positive as they are percentages.

To account for the limited sample size, we tested the model with 1000 different seeds for more robust and reliable scores. Furthermore, we integrated Bootstrap .632+ as an internal validation technique. This method provides information on how the model may behave in different datasets not involved in the training phase<sup>69</sup>. Additionally, we repeatedly run the model, removing one of the five datasets that have

RNA-Seq data respectively. The AUC did not change by more than 0.016 (2.14%) when UNC-108 was removed, and by 0.007 (0.94%) for the rest of the datasets, ensuring that our results are not driven by one of the datasets.

Finally, we tested the model in the independent validation cohort JAVELIN Bladder 100 trial. We were provided with processed gene expression and mutation call data. From the TPM matrices, we obtained the gene expression levels of selected genes (PD-1, PD-L1, CCND1), immune signatures using GSVA (IFN- $\gamma$ , stroma/EMT, inflamed T cells, TGF- $\beta$ , and antigen-presenting machinery pathways), and immune cell abundances from CIBERSORT (M1 macrophages, CD4 memory activated T cells, CD8 T cells, and regulatory T cells). From the list of called mutations per patient, we obtain an estimate of TMB. Due to the lack of raw WES data and incomplete clinical information, we could not determine the number of non-stop mutations, the APOBEC-enrichment score, the status of liver metastasis, or the ECOG score of the patients. Moreover, raw RNA-Seq data would have been needed to compute the signature of tumor-specific lncRNAs. We used the TMB + RNA model (Fig. 6A), the lncRNA signature was given a 0 value.

### Statistical analysis

Plots were mainly generated using R (version 4.1.2), Rstudio (version 1.4) and Python (version 3.8.6). When comparing continuous variables between the two response groups, significance levels were obtained using the two-sided, two-sample Wilcoxon signed-rank test. We used the Bonferroni correction post-hoc to assess the significance of the effect of expression of single genes or gene signatures. As 15 gene signatures were tested, the adjusted alpha for gene signatures was 0.00033. For genes, the alpha for adjusted  $p$  values was 0.00029. When comparing ratios the Fisher exact test was employed. Pearson correlation was used to compare numeric variables.

### Reporting summary

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

### Data availability

The Supplementary Information file contains Supplementary Figures and Tables indicated in the text. Supplementary Data 1 file contains a list of the variables investigated, descriptive statistics of the distribution of the variable values in different groups of patients, and numerical values related to the plots in the Figures. Source data are provided as a Source Data file with this paper. The clinical data analyzed in this study can be accessed in the following databases: IMvigor210 (EGA study ID EGAS00001002556)<sup>9</sup>, HdM-BLCA-1 (EGA study ID EGAS00001007086)<sup>18</sup>, SNY-2017 (dbGaP accession ID phs001743)<sup>5</sup>, MIAO-2018 (dbGaP accession ID phs001565)<sup>3</sup>, UC-GENOME (dbGaP accession ID phs003066)<sup>19</sup>, UNC-108 (Omnibus accession ID GSE176307)<sup>16</sup> and the JAVELIN Bladder 100 (ClinicalTrials.gov identifier: NCT026034329)<sup>45</sup>. Source data are provided with this paper.

### Code availability

The code for the model and decision trees is available from the github repository [https://github.com/EvolutionaryGenomics-GRIB/ML\\_Pipeline](https://github.com/EvolutionaryGenomics-GRIB/ML_Pipeline). The trained model can be executed at [https://github.com/EvolutionaryGenomics-GRIB/BLCA\\_ICI\\_Response\\_Predictor](https://github.com/EvolutionaryGenomics-GRIB/BLCA_ICI_Response_Predictor) (release v.1.0.0)<sup>70</sup>.

### References

1. Bellmunt, J. et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N. Engl. J. Med.* **376**, 1015–1026 (2017).
2. Powles, T. et al. Avelumab maintenance therapy for advanced or metastatic urothelial carcinoma. *N. Engl. J. Med.* **383**, 1218–1230 (2020).
3. Miao, D. et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat. Genet.* **50**, 1271–1281 (2018).
4. Nadal, R., Valderrama, B. P. & Bellmunt, J. Progress in systemic therapy for advanced-stage urothelial carcinoma. *Nat. Rev. Clin. Oncol.* **21**, 8–27 (2024).
5. Snyder, A. et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. *PLoS Med.* **14**, e1002309 (2017).
6. McGrail, D. J. et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* **32**, 661–672 (2021).
7. Davis, E. J. et al. Clinical Correlates of Response to Anti-PD-1-based therapy in patients with metastatic melanoma. *J. Immunother.* **42**, 221–227 (2019).
8. Galsky, M. D. et al. Perioperative pembrolizumab therapy in muscle-invasive bladder cancer: Phase III KEYNOTE-866 and KEYNOTE-905/EV-303. *Futur. Oncol.* **17**, 3137–3150 (2021).
9. Mariathasan, S. et al. TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
10. Robertson, A. G. et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017).
11. Kamoun, A. et al. A consensus molecular classification of muscle-invasive bladder cancer. *Eur. Urol.* **77**, 420–433 (2020).
12. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
13. Rosenberg, J. E. et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909–1920 (2016).
14. Kim, J. et al. The Cancer Genome Atlas Expression subtypes stratify response to checkpoint inhibition in advanced urothelial cancer and identify a subset of patients with high survival probability. *Eur. Urol.* **75**, 961–964 (2019).
15. Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).
16. Damrauer, J. S. et al. Collaborative study from the Bladder Cancer Advocacy Network for the genomic analysis of metastatic urothelial cancer. *Nat. Commun.* **13**, 6658 (2022).
17. Long, J. et al. A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. *Genome Med.* **14**, 20 (2022).
18. Boll, L. M. et al. The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer. *Sci. Rep.* **13**, 15287 (2023).
19. Rose, T. L. et al. Fibroblast growth factor receptor 3 alterations and response to immune checkpoint inhibition in metastatic urothelial cancer: a real world experience. *Br. J. Cancer* **125**, 1251–1260 (2021).
20. Martínez-Jiménez, F. et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
21. Yang, Y. et al. Targeting ARHGEF12 promotes neuroblastoma differentiation, MYCN degradation, and reduces tumorigenicity. *Cell. Oncol.* **46**, 133–143 (2023).
22. Galsky, M. D. et al. Impact of tumor mutation burden on nivolumab efficacy in second-line urothelial carcinoma patients: Exploratory analysis of the phase II checkmate 275 study. *Ann. Oncol.* **28**, v296–v297 (2017).
23. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and



- integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
24. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
  25. Nassar, A. H. et al. A model combining clinical and genomic factors to predict response to PD-1/PD-L1 blockade in advanced urothelial carcinoma. *Br. J. Cancer* **122**, 555–563 (2020).
  26. Adib, E. et al. *CDKN2A* Alterations and Response to Immunotherapy in Solid Tumors. *Clin. Cancer Res.* **27**, 4025–4035 (2021).
  27. Steen, C. B., Liu, C. L., Alizadeh, A. A. & Newman, A. M. *Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx*. p. 135–157 (Humana, New York, NY, 2020).
  28. Philips, G. K. & Atkins, M. Therapeutic uses of anti-PD-1 and anti-PD-L1 antibodies. *Int. Immunol.* **27**, 39–46 (2015).
  29. Büttner, R. et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. *J. Clin. Oncol.* **35**, 3867–3876 (2017).
  30. Baxi, V. et al. Association of artificial intelligence-powered and manual quantification of programmed death-ligand 1 (PD-L1) expression with outcomes in patients treated with nivolumab ± ipilimumab. *Mod. Pathol.* **35**, 1529–1539 (2022).
  31. Shen, X. & Zhao, B. Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: meta-analysis. *BMJ* **362**, k3529 (2018).
  32. McDermott, D. F. et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat. Med.* **24**, 749–757 (2018).
  33. Fehrenbacher, L. et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837–1846 (2016).
  34. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
  35. Ayers, M. et al. IFN- $\gamma$ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Investig.* **127**, 2930–2940 (2017).
  36. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593 (2018).
  37. Hernando-Calvo, A. et al. A pan-cancer clinical platform to predict immunotherapy outcomes and prioritize immuno-oncology combinations in early-phase trials. *Med* **4**, 710–727.e5 (2023).
  38. Benguigui, M. et al. Interferon-stimulated neutrophils as a predictor of immunotherapy response. *Cancer Cell* **42**, 253–265.e12 (2024).
  39. Thompson, J. C. et al. Gene signature of antigen processing and presentation machinery predicts response to checkpoint blockade in non-small cell lung cancer (NSCLC) and melanoma. *J. Immunother.* **8**, e000974 (2020).
  40. Ouspenskaia, T. et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01021-3> (2021).
  41. Ruiz Cuevas, M. V. et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep* **34**, 108815 (2021).
  42. Laumont, C. M. et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, eaau5516 (2018).
  43. Wang, L. et al. EMT- and stroma-related gene expression and resistance to PD-1 blockade in urothelial cancer. *Nat. Commun.* **9**, 3503 (2018).
  44. Sha, D. et al. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* **10**, 1808–1825 (2020).
  45. Powles, T. et al. Avelumab maintenance in advanced urothelial carcinoma: biomarker analysis of the phase 3 JAVELIN Bladder 100 trial. *Nat. Med.* **27**, 2200–2211 (2021).
  46. Togashi, Y., Shitara, K. & Nishikawa, H. Regulatory T cells in cancer immunosuppression — implications for anticancer therapy. *Nat. Rev. Clin. Oncol.* **16**, 356–371 (2019).
  47. Boichard, A. et al. APOBEC-related mutagenesis and neo-peptide hydrophobicity: implications for response to immunotherapy. *Oncoimmunology* **8**, 1550341 (2019).
  48. Bellmunt, J. et al. Genomic predictors of good outcome, recurrence, or progression in high-grade T1 non-muscle-invasive bladder cancer. *Cancer Res.* **80**, 4476–4486 (2020).
  49. Turajlic, S. et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet. Oncol.* **18**, 1009–1021 (2017).
  50. Zhang, J. et al. Cyclin D–CDK4 kinase destabilizes PD-L1 via cullin 3–SPOP to control cancer immune surveillance. *Nature* **553**, 91–95 (2018).
  51. Barczak, W. et al. Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nat. Commun.* **14**, 1078 (2023).
  52. Xiong, D. et al. Immunogenomic landscape contributes to hyper-progressive disease after anti-PD-1 immunotherapy for cancer. *iScience* **9**, 258–277 (2018).
  53. Chowell, D. et al. Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nat. Biotechnol.* **40**, 499–506 (2022).
  54. Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).
  55. Griffin, J. et al. Verification of molecular subtyping of bladder cancer in the GUSTO clinical trial. *J. Pathol. Clin. Res.* **10**, e12363 (2024).
  56. Balar, A. V. & Weber, J. S. PD-1 and PD-L1 antibodies in cancer: current status and future directions. *Cancer Immunol. Immunother.* **66**, 551–564 (2017).
  57. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
  58. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
  59. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
  60. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
  61. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
  62. Jurtz, V. et al. NetMHCpan-4.0: improved peptide-MHC Class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
  63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
  64. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
  65. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
  66. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
  67. Tokunaga, R. et al. 12-Chemokine signature, a predictor of tumor recurrence in colorectal cancer. *Int. J. Cancer* **147**, 532–541 (2020).

68. Chong, C. et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* **11**, 1293 (2020).
69. Efron, B. & Tibshirani, R. Improvements on cross-validation: the 632+ Bootstrap Method. *J. Am. Stat. Assoc.* **92**, 548–560 (1997).
70. Boll, L. M., & Vázquez Montes de Oca, S. BLCA\_ICI\_Response\_Predictor (Version 1.0.0). GitHub repository. [https://github.com/EvolutionaryGenomics-GRIB/BLCA\\_ICI\\_Response\\_Predictor](https://github.com/EvolutionaryGenomics-GRIB/BLCA_ICI_Response_Predictor), 10.5281/zenodo.14095067 (2024).

## Acknowledgements

The work was supported by the following grants and agencies: 1. Research project PID2019-105595GB-I00 (R.C.) funded by MICIU/AEI/10.13039/501100011033 and PID2021-122726NB-I00 (MMA) funded by MCIN/AEI/10.13039/501100011033 and by “ERDF: A way of making Europe”, by the “European Union”; 2. Project PI22/00171 (J.B.), funded by Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union ; 3. 2021SGR00042 by Generalitat de Catalunya (MMA); 4. Ayudas Fundación BBVA a Proyectos de Investigación Científica en Biomedicina 2021 (MMA). LMB is funded by an INPhINIT PhD fellowship from “la Caixa” Foundation (ID 100010434), under the agreement LCF/BQ/DI21/11860060.

## Author contributions

L.M.B. and S.V.M. contributed equally to this work. J.P.B., M.M.A., S.V.M. and L.M.B. gathered and integrated the datasets of this study. M.E.C., S.V.M. and L.M.B. performed the computational analysis of RNA-seq data. L.M.B. conducted the computational analysis of WES data. S.V.M. and R.C. developed the machine learning models. M.M.A., J.P.B. and J.B. contributed to the conceptualization of the study and design of experiments. M.M.A., L.M.B., S.V.M., and M.E.C. wrote the manuscript with contributions from all co-authors.

## Competing interests

Potential conflicts of interest: J.B. has served in consulting or advisory roles for Astellas Pharma, AstraZeneca/MedImmune, Bristol Myers Squibb, Genentech, Novartis, Pfizer, Pierre Fabre, and the healthcare business of Merck KGaA, Darmstadt, Germany; has received travel and accommodation expenses from Ipsen, Merck & Co., Kenilworth, NJ, and Pfizer; reports patents, royalties, other intellectual property from UpToDate; reports stock and other ownership interests in Rainier

Therapeutics; has received honoraria from UpToDate; and has received institutional research funding from Millennium, Pfizer, Sanofi, and the healthcare business of Merck KGaA, Darmstadt, Germany. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56462-0>.

**Correspondence** and requests for materials should be addressed to Joaquim Bellmunt, Júlia Perera-Bel or M. Mar Albà.

**Peer review information** *Nature Communications* thanks Mark Linch, Ludmila Prokunina-Olsson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

### 3.3 Stop-loss mutations in cancer

**Authors:** Lilian M. Boll, Marta E. Camarena, Joan A. Martorell, Júlia Perera-Bel, M. Mar Albà

**Status:** Manuscript in preparation

**Supplementary Material:** Appendix, section 5.3.

#### Abstract

Stop-loss mutations extend protein translation into the 3' untranslated region (3' UTR), with potential consequences for tumorigenesis and immune recognition. These mutations have been linked to immunotherapy response and protein loss in tumor suppressor genes. In this study, we analyzed a dataset of 2,066 stop-loss mutations across 32 cancer types to identify recurrently mutated genes. Among several recurrently mutated genes, we observed the same stop-loss mutation in the oncogene *PTMA* 14 times, highlighting a potential functional relevance for cancer progression. Additionally, we found that oncogenes were significantly enriched in stop-loss mutations, whereas tumor suppressor genes were not. To explore the potential immunogenicity of these mutations, we examined immunopeptidomics data and identified peptides corresponding to stop-loss-derived extensions. These findings suggest that stop-loss mutations could contribute to tumorigenesis and immune recognition, offering new insights into their role in cancer biology and immunotherapy. Further research is needed to explore their functional impact and therapeutic potential.

## Instructions

Stop codons (UAA, UAG, and UGA) terminate the ribosomal translation of a coding sequence. A mutation converting a stop codon to a sense codon, referred to as a stop-loss, stop-lost, or non-stop mutation, causes translation into the 3' untranslated region (UTR). The extension of the protein continues up to the next in-frame stop codon or the polyA tail of the mRNA molecule. A bioinformatics analysis by Shibata and Ohaka *et al.* (2015) showed that most canonical stop codons are followed by a downstream stop codon, thus cases, where the mRNA lacks a stop codon, are rare, and their translation is expected to be limited by destabilizing and accelerated degradation of the mRNA molecule and the nascent protein (Klauer and van Hoof, 2012; Shibata *et al.*, 2015). The same study reported that the average extension of the peptide sequence is 27.8 amino acids, while another study reports a median extension of 18 amino acids (Dhamija *et al.*, 2020). Similarly, Flores *et al.* report that around 5% of the stop-loss mutations were missing a downstream stop codon (Flores *et al.*, 2024).

Around 0.2% of the non-synonymous mutations in the Human Gene Mutation Database are stop-loss mutations. Some of these have been related to heritable diseases, underlining the effect of the mutation on the functionality of the protein (Shibata *et al.*, 2015). The translation into the 3'UTR can also be caused by a readthrough where the ribosome does not terminate at the stop codon (Arribere *et al.*, 2016). The resulting C-terminal extension can be digested into peptides and loaded onto a MHC in the endoplasmic reticulum.

Stop-loss mutations have also been observed in cancer. A well-known example is a stop-loss mutation generating an extension of 40 amino acids in the tumor suppressor *SMAD4* (Dhamija *et al.*, 2020). This extension is associated with increased degradation of the protein via a newly created hydrophobic degron sequence. Other examples of protein loss due to stop-loss mutations are the renal tumor suppressor genes *BAP1*, *PTEN*, and *VHL* (Pal *et al.*, 2025). While the short

stop-loss extension of *PTEN* and *VHL* are thought to affect the proteins' stability, leading to proteasomal degradation, stop-loss mutations in *BAP1* result in a very long extension that leads to translation inhibition of the mRNA. A study in which the effect of 2,335 cancer stop-loss mutations in protein levels was tested indicated that protein degradation-promoting extensions tend to have higher hydrophobicity than those that do not have such an effect (Ghosh et al., 2024).

In a previous study, we reported that a high stop-loss mutational burden is significantly associated with a positive response to immunotherapy in bladder cancer (BLCA) patients (Boll et al., 2025). One possibility is that some of the stop-loss mutations create tumor-specific neoantigens that evoke an immune response against the cancer cells. The potential immunogenic effect of peptides originating from the readthrough of a stop codon is also supported by the work of Goodenough and colleagues where aminoglycosides, a drug against premature stop codon mutations, was found to enhance stop codon readthrough, leading to an increased CD8+ T cell response due to HLA-I presented epitopes (Goodenough et al., 2014).

To better understand the impact of stop-loss mutations in cancer, we have performed an analysis of the mutational patterns in tumor samples from 12,224 cancer patients. This has increased our initial set of 85 stop-loss mutations in BLCA (Boll et al., 2025) to 2,066 stop-loss in 32 different cancer types. Using this data, we have identified several new recurrent stop-loss mutations, which might have cancer-promoting effects. We have also observed that the proportion of stop-loss mutations versus missense mutations shows a significant enrichment in oncogenes when compared to tumor suppressor genes, supporting the oncogenicity of some of these mutations. Finally, we have obtained evidence that stop-loss mutations can generate novel tumor-specific antigens, which could be useful for therapeutic applications.

## Results

### 3.3.1 Building a catalogue of cancer-associated stop-loss mutations

We have built a catalogue of 2,066 stop-loss mutations found in samples from different cancer patient cohorts (Table 3.1, Table S1). The mutations were found in 1,284 patient samples out of the 12,224 analyzed (10.5%), and they were located in 1,821 different protein-coding genes. Most stop-loss mutations were identified in tumors of the genitourinary system (980), followed by the respiratory system (424), and digestive system (380). As expected, cancer types that have been reported to have a high mutational load also carried many stop-loss mutations; this included bladder cancer (253), lung (LUAD 187 and LUSC 157), and melanoma (168). The highest number of stop-loss mutations was derived from patients with uterine corpus endometrial carcinoma (423).

Mutation recurrence is a hallmark of cancer-driver mutations. We identified 179 recurrently stop-loss mutated genes, present in the samples of at least two patients (9.83% of the stop-loss mutations). Several genes had stop-loss mutations in several cancer samples, including 36 genes that were mutated in at least three patients (Table 3.2). One of these genes, *PTMA*, which encodes the protein polypeptide prothymosin alpha (proT $\alpha$ ), was mutated in 14 different samples. Eight other genes, including cyclin T1 (*CCNT1*) and immunoglobulin kappa constant (*IGKC*), were mutated between 4 and 9 times.

We also investigated if stop-loss mutations tended to be more clonal than missense mutations. However, we found no significant differences in the proportion of clonal versus subclonal mutations between the two types of mutations (Table S2).

**Table 3.1: Overview of the datasets included in this study.** Cohort names, cancer type, published data accession number, and the accessed data type. Matched WES indicates that the mutations were called using paired tumor/control samples. VCF file indicates that the mutations were extracted from the variant calling files from the corresponding studies. BLCA: bladder cancer; HNSC: head and neck squamous cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma, NSCLC: non-small cell lung cancer; SKCM: skin cutaneous melanoma. \*The number of TCGA patients and cancer types analyzed can be found in Table S1

Dataset	N	Cancer Type	Accession Portal	Accession number	Mutation source	Non-syn. mutations	Stop-loss mutations (%)
IMvigor210	240	BLCA	EGA	EGAS00001002556	Matched WES	48696	68 (0.14)
HdM-BLCA-1	27	BLCA	EGA	EGAS00001007086	Matched WES	5781	6 (0.10)
MIAO-2018	27	BLCA	dbGaP	phs001565	Matched WES	5269	6 (0.11)
SNY-2017	25	BLCA	dbGaP	phs001743	Matched WES	4477	5 (0.11)
FIS_20200101	97	BLCA	medRxiv	-	Tumor WES	16657	24 (0.14)
Chong	9	LUAD, SKCM	EGA	EGAS00001003723	Matched WES	4509	4 (0.09)
Kraemer	4	SKCM	EGA	EGAS00001006298	Matched WES	889	1 (0.11)
Bassani-Sternberg	5	SKCM	EGA	EGAS00001002050	VCF file	6856	1 (0.01)
Loeffler	17	LIHC	Figshare	-	VCF file	2250	3 (0.04)
Roper/Yue	15	LUAD	dbGaP	phs002001	VCF file	5368	6 (0.11)
Samstein	1611	NSCLC, SKCM, HNSC	cBioPortal	tmb_mskcc_2018	VCF file	20032	6 (0.03)
TCGA*	10147	33 cancer types	GDC Portal	-	VCF file	2147998	1936 (0.09)
<b>Total</b>	<b>12224</b>					<b>2268782</b>	<b>2066 (0.09)</b>

## Stop-loss mutations in the *PTMA* gene

The most recurrently mutated gene was prothymosin alpha (PTMA), which had a stop-loss mutation in 14 TCGA patients. Interestingly, all patients had the same T>C polymorphism in the TAG stop codon leading to a translation of a glutamine (CAG → Q) and an extension of nine amino acids to the next in-frame downstream stop codon (Figure 3.1). None of the other seven possible changes leading to other amino acids was observed (Figure S1). The mutation was found in 7 samples of testicular germ cell cancer (TGCT), 1 colorectal cancer (COAD), 1 cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), 1 esophageal carcinoma (ESCA), 1 kidney renal clear cell carcinoma (KIRC), 1 liver hepatocellular carcinoma (LIHC), 1 sarcoma (SARC) and 1 uterine corpus endometrial carcinoma (UCEC). Additionally, the same mutation was found in two cancer patients in the nonstop

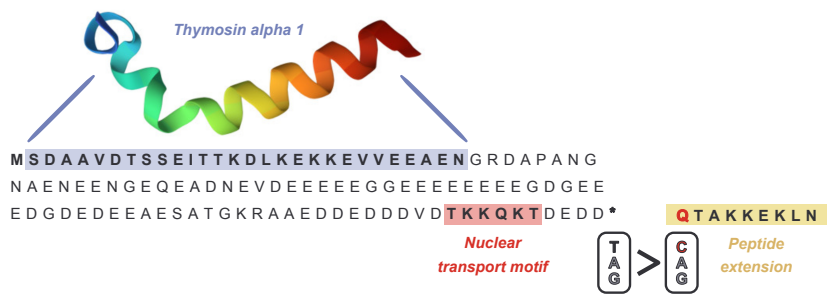
**Table 3.2: Most recurrent stop-loss mutations over all datasets.** Listed are the 36 genes with stop-loss mutations found in at least three patients.

Gene	Stop-loss mutation	Tumor source	System	Cancer types
<i>PTMA</i>	14	TCGA	digestive, genitourinary, neurological	TGCT, CESC, COAD, ESCA, KIRC, LIHC, SARC, UCEC
<i>CSN2</i>	9	TCGA	genitourinary, neurological	GBM, UCEC
<i>CCNT1</i>	5	TCGA	cardiovascular, genitourinary	LAML, OV
<i>IGKC</i>	5	TCGA	digestive, genitourinary	BLCA, BRCA, CESC, COAD
<i>RDH16</i>	5	IMvigor210, TCGA	genitourinary, integumentary, respiratory	BLCA, MESO, SKCM
<i>ST6GALNAC3</i>	5	TCGA	respiratory	HNSC, LUAD
<i>CPS1</i>	4	TCGA	respiratory	LUAD
<i>LARP4</i>	4	TCGA	cardiovascular, genitourinary	LAML, OV
<i>PCDH9</i>	4	TCGA	digestive, genitourinary	BLCA, COAD, ESCA, STAD
<i>AC026703.1</i>	3	TCGA	digestive, genitourinary	COAD, READ, UCEC
<i>BCL7C</i>	3	TCGA	digestive, genitourinary	CECSC, STAD, UCEC
<i>CADM1</i>	3	TCGA	digestive, genitourinary	LIHC, PRAD
<i>CCL7</i>	3	TCGA	genitourinary, respiratory	LUAD, LUSC, UCEC
<i>CGN</i>	3	TCGA	integumentary, neurological, respiratory	LUSC, SARC, SKCM
<i>CLEC2B</i>	3	IMvigor210, TCGA	digestive, genitourinary	BLCA, COAD, STAD
<i>CLSTN2</i>	3	TCGA	respiratory	LUAD
<i>DEFB119</i>	3	TCGA	genitourinary	UCEC
<i>EDDM3A</i>	3	TCGA	respiratory	LUAD
<i>EXTL2</i>	3	TCGA	cardiovascular, genitourinary	LAML, OV
<i>FANCI</i>	3	TCGA	digestive, neurological	ESCA, GBM
<i>HIST1H4K</i>	3	Roper, TCGA	digestive, genitourinary, respiratory	CECSC, ESCA, LUAD
<i>HIST1H4L</i>	3	TCGA	digestive, genitourinary	BLCA, STAD
<i>HSPA8</i>	3	TCGA	digestive	COAD
<i>IDH3B</i>	3	TCGA	genitourinary, lymphatic	KIRP, THYM
<i>MPP2</i>	3	TCGA	genitourinary, integumentary, respiratory	LUAD, SKCM, UCEC
<i>OR52A5</i>	3	TCGA	genitourinary	KIRC
<i>OR5D16</i>	3	TCGA	genitourinary	UCEC
<i>PADI1</i>	3	TCGA	digestive	COAD
<i>PLA2G1B</i>	3	TCGA	digestive, genitourinary, respiratory	BLCA, HNSC, READ
<i>PLCB1</i>	3	TCGA	respiratory	LUAD
<i>PRDM4</i>	3	TCGA	cardiovascular, neurological	GBM, LAML
<i>SF3B3</i>	3	TCGA	integumentary, respiratory	LUAD, SKCM
<i>SKIL</i>	3	TCGA	genitourinary	BRCA, UCEC
<i>TMTC3</i>	3	TCGA	respiratory	LUAD
<i>ZIC5</i>	3	TCGA	digestive, genitourinary	COAD, UCEC
<i>ZIM3</i>	3	TCGA	respiratory	LUAD



database, in TGCT and ganglioblastoma (<https://nonstopdb.dkfz.de/>, accessed 31<sup>st</sup> of Oct 2024) (Dhamija et al., 2020).

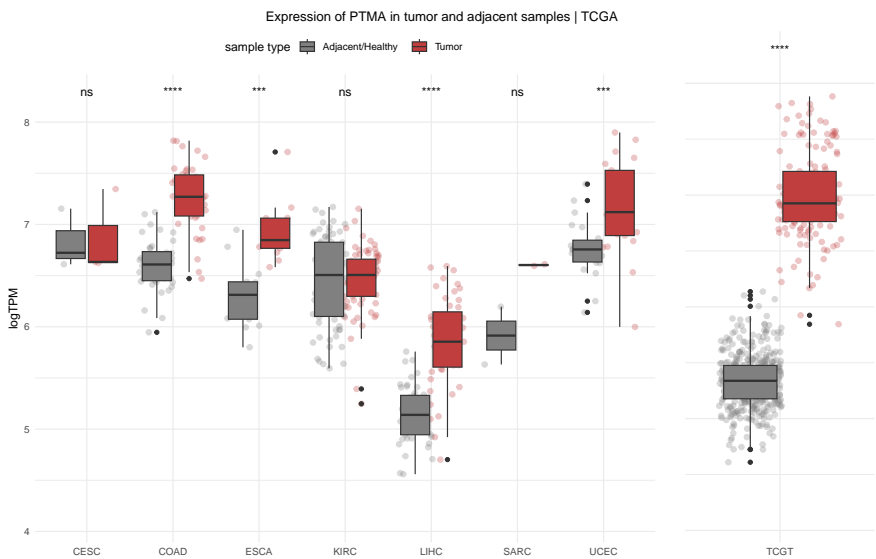
*PTMA* encodes the polypeptide prothymosin alpha (proT $\alpha$ ), a relatively small protein with a length of 111 amino acids and a large proportion of negatively charged amino acids (39 glutamic (E) and 18 aspartic acids (D)) (Figure 3.1). Both the protein sequence as well as the C-terminal extension ('QTAKKEKLN') are hydrophilic (GRAVY hydrophobicity index of -1.97 and -1.92, respectively). Predicting the HLA I binding probability of peptides from the protein extension did not yield potential neoantigens. The protein has a nuclear transport motif, TKKQKT, at the protein's C-terminus (Manrow et al., 1991) (Figure 3.1).



**Figure 3.1: Sequence of the proT $\alpha$  protein. The extension generated by the T>C mutation in the stop codon is indicated.** In the thymus, the protein is cleaved and thymosin alpha 1 is produced (position 2-29). The 3D structure of thymosin alpha is taken from UniProt P06454.

The protein proT $\alpha$  is considered an oncoprotein and a biological response modifier with effects on the immune system (Samara et al., 2016). Intracellularly, it affects proliferation and cell survival (Enkemann et al., 2000) and prevents apoptosis by inhibiting apoptosome formation (Jiang et al., 2003; Malicet et al., 2006). The transcription of *PTMA* increases in proliferating cells, as well as its post-translational phosphorylation and its presence in the nucleus, where it supports chromatin decondensation by histone binding of the central domain (Pérez-Estévez et al., 1997; Segade and Gómez-Márquez, 1999; Trumbore et al., 1997). In the thymus, proT $\alpha$  is cleaved generating thymosin alpha, which has anti-tumoral activity by stimulating T cell production.

Examination of *PTMA* expression in multiple tissues indicates that it is expressed ubiquitously (Figure S2). The highest gene expression is observed in EBV-transformed lymphocytes with a median of 1,757 TPM. The lowest expression is in muscle, testis cells, blood, and heart (all median under 200 TPMs). In accordance with early findings showing an increased expression level for *PTMA* in malignant tissue compared to healthy tissue (Tsitsiloni et al., 1994), we observed that *PTMA* tends to be overexpressed in cancer tissues compared to matched controls (Figure 3.2). The trend is very strong in tissues with recurrent *PTMA* stop-loss mutations, such as TCGT and COAD (Figure 3.2). In patients where mutation and expression data is available, the *PTMA* stop-loss mutation does not seem to affect the transcript levels (Figure S3).



**Figure 3.2: Expression levels of *PTMA* in cancer and matched controls.** Comparison of the *PTMA* gene expression values for matched cancer/control data, for different cancer types (left). CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; KIRC: kidney renal clear cell carcinoma; LIHC: liver hepatocellular carcinoma; SARC: sarcoma; UCEC: uterine corpus endometrial; TCGT: testicular germ cell tumors. For TCGT, healthy testis expression from the GTEx project is shown compared to the expression in tumor tissue, as no paired data was available. Each dot represents a patient. The line in the box plot indicates the median value. The values were compared using a Wilcoxon test. ns: no significant; \*\*\*: p-value  $\leq 0.001$ ; \*\*\*\*: p-value  $\leq 0.0001$ .

## Oncogenes are enriched in stop-loss mutations versus missense mutations

Similarly to previous observations for missense mutations, stop-loss mutations were more prevalent in genes already known to have roles in cancer than in other genes. We found 26 stop-loss mutations in a set of 77 cancer-associated genes previously described in the literature (Robertson et al. (2017), Table S3), a higher proportion than the approximately 1 stop-loss mutation per 10 genes observed in general. These genes contained 1.26% of all stop-loss mutations and 0.73% of all missense mutations, the first type being significantly overrepresented over the second one (X2-test p-value=0.0045). We then examined the mutations in lists of oncogenes and tumor suppressor genes from COSMIC (Table S3). We found that only the first class showed significant enrichment of the proportion of stop loss mutations versus missense mutations (oncogenes n=319, 2.13% versus 1.37%, respectively, X2-test p-value=0.0029; tumor suppressor genes n=320, 1.93% versus 1.60%, respectively X2-test p-value=0.25). Stop-loss mutations have been related to loss of protein expression and tumor suppressor gene inactivation by increased protein degradation, such as in the cases of *SMAD4*, *PTEN* and *VHL* (Dhamija et al., 2020; Pal et al., 2025). Our results indicate that they could also be important in promoting the functions of oncogenes in the context of cancer.

Several of the cancer-related genes showed recurrent mutations. The transmembrane mucin gene *MUC4* has been related to aggressive behavior and poor outcome, especially in epithelial cancers (Xia et al., 2016). We found two different stop-loss mutations in this gene in two female patients with cervical carcinoma (CESC) and uterine corpus endometrial carcinoma (UCEC). *MUC4* is involved in epithelial differentiation and renewal and has been suggested as a marker for cervical cancer tissue (Munro et al., 2009). We further found two other patients with colon cancer (COAD) and uterine corpus endometrial carcinoma (UCEC) carry a stop-loss mutation in *RPL22*. Notably, deleterious missense and frameshift mutations in this gene have been

repeatedly reported in microsatellite-unstable endometrial and colorectal carcinoma (Ferreira et al., 2014; Novetsky et al., 2013; Weinstein et al., 2024). Another example was the gene *BCL11B*, mutated in two patients with glioblastoma (GLM). The transcription factor is involved in several pathways but mainly affects the development of the nervous tissue, as well as T cell differentiation (García-Aznar et al., 2024). Due to its pivotal role, the gene can be considered oncogenic or tumor-suppressive. Finally, two patients with cervical (CESC) and skin cell melanoma (SKCM) carried a stop-loss mutation in *SDHD*. The gene is involved in the citric cell cycle and oxidative phosphorylation. Two studies have investigated the negative effects of SDH expression loss due to mutations in the promoter regions of *SDHD* in melanoma patients (Scholz et al., 2015; Weinhold et al., 2014).

### Highly recurrently mutated genes

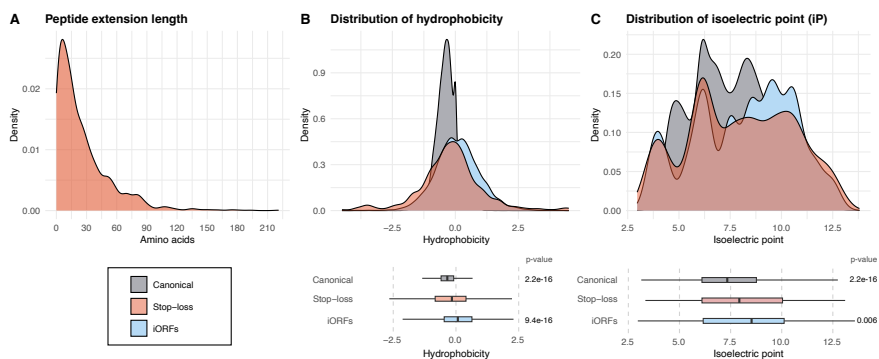
Other recurrently mutated genes were the cyclin *CCNT1* and *RDH16*, involved in the metabolism of retinoids, both found to be mutated in five patients. Further, we found *IGKC*, *LARP4*, and *PCDH9* to be stop-loss mutated in four patients. *LARP4* is involved in the RNA metabolism, stabilizing the poly-A tails, and was found to be mutated in 3 patients with ovarian cancer and a patient with leukemia. The immunoglobulin kappa constant *IGKC* affects the immune system by enabling antigen binding activity and immunoglobulin receptor activity. An additional 18 genes were mutated in three cancer patients. Among them *HIST1H4K* and *HIST1H4L*, which are both part of the histone 4 family crucial during cell replication. Other stop-loss mutations were found for histone 1 (*HIST1H1C*, *HIST1H1E*), histone 2 (*HIST1H2AI*, *HIST1H2BD*, *HIST1H2BJ*, *HIST1H2BN*), and histone 3 (*HIST1H3C*, *HIST1H3E*). All of these stop-loss mutations are located in the H1 cluster, the largest histone gene cluster located at 6p22.2 containing more than 50 histone genes.

## Peptide extensions from stop-loss have similar properties as potential intronic ORFs

Next, we computed 1066 unique protein extensions resulting from stop-loss mutations. Cases with no downstream stop codon found in the transcript were discarded. The median length of an extension was 16 amino acids, with 95% being shorter than 80 amino acids (Figure 3.3). This value is similar to the average extension of 18 amino acids predicted for the stop-loss mutations in the NonStopDB (Dhamija et al., 2020). We were further interested in testing if the extensions had different amino acid composition and properties than canonical sequences and intronic regions. For this, we translated potential intronic open reading frames (iORFs) in human introns. We focused on hydrophobicity and isoelectric point. While the stop-loss extensions and iORFs showed a similar, symmetrical hydrophobicity distribution, canonical sequences were overall less hydrophobic with a much more compact distribution ( $p\text{-value} = 2.2\text{e-}16$ ). Further, most canonical proteins had an isoelectric point (iP) under 10 (90%), with several local maxima. Again, the stop-loss extensions resemble more the pattern of intronicORFs than the canonical sequences, being less centered around the mean body pH of 7.35-7.45, and showing an extended distribution towards more basic values.

## HLA I-bound peptides generated by stop-loss mutations

The protein extensions resulting from the stop-loss mutations can be a source of tumor-specific antigens. Analysis of two immunopeptidomics datasets from melanoma comprising 30 patients (Bassani-Sternberg et al., 2016; Chong et al., 2020) identified 10 tumor-specific peptides that corresponded to extensions derived from stop-loss mutations (Table 3.3). In three cases the mutations had originally also been identified in melanoma samples (SKCM), indicating that they could be recurrent cases. We identified repeated hits to an extension in the gene *NME1*, which codes for nucleoside diphosphate kinase (NDK). The gene is expressed at reduced levels in some highly metastatic



**Figure 3.3: Properties of extensions from stop-loss mutations compared to canonical proteins and intronic ORFs.** **A:** Amino acid length of predicted peptide extension resulting from stop-loss mutations (N=1,590). **B:** Hydrophobicity in stop-loss extensions, canonical peptides, and intronic open reading frames (iORFs). Hydrophobicity was computed using the KyteDoolittle scale. **C:** Isoelectric point in the stop-loss extensions, canonical peptides, and iORFs. The isoelectric point was calculated using the EMBOSS pK scale. Wilcoxon test was used to compare stop-loss extensions to iORFs and canonical sequences. The statistics of the distributions are provided in Table ??.

cells, but increased *NME1* levels are found in neuroblastoma. The findings support our hypothesis that stop-loss mutations can generate new tumor-specific antigens.

**Table 3.3: Peptide hits of mass spec immunopeptidomics data from melanoma patient samples.** We show the significant hits to all possible peptides newly generated by the stop-loss mutations detected in the different cancer samples. The first five different peptides are fully located in the extension; the other five span the protein C-terminus and part of the extended region. The hits were identified with MHCquant (FDR < 1%). The source refers to the immunopeptidomics dataset in which hits were detected. PXD004894 consists of 23 patients and PXD013649 consists of 7 patients. Tumor refers to the cancer type in which the mutations were originally detected.

Peptide	Gene	Source	Samples	Tumor	Protein Extension
<i>EVSPAPWNR</i>	<i>RGS14</i>	PXD013649	T1185B	KIRP	QLPNSPGQLHGTRRAEHAMGPLCMPCLCHECP WPLPAMGRPAGRAGRGGKGTQMRHTPQLPPPC PSTSSPPIPCSQATMGEVSPAPWNR <del>LA</del> QHGGMA ALAVPASACAKLQQGQEEGPAPPQEAGMSKALR VQAGSPVPHPHRIYCTYRFC SRLGAAGFVLDV
<i>TAAVKKHAL</i>	<i>IL1RAPL1</i>	PXD004894	MM4	SKCM	QKSKGHPVPGRLSGICSPVPGTKSS <del>TA</del> AVKKHAL ESLEHEEKQGLVHMFFGISL
<i>RRSILDHLIT</i>	<i>CIAO1</i>	PXD004894	MM8	SKCM	ATSTLDRVMT <del>PQ</del> KTSYKTL <del>PAPER</del> RRSILDHLIT WLTSLQTWVEVQSHRI <del>AF</del> LPRL
<i>YLQELHHNL</i>	<i>NME1</i>	PXD004894 (1), PXD013649 (3)	MM33, Me275, Me290, T1015A	UCEC	QEGRPHCF <del>SH</del> PFLLPMGRGPGCRKSSYLQELHH NLEGSSWSCEFSLYSVTIPDHLIKMLPPSIGFIELVT SYCCIAFFSFFSCTLNNLT
<i>ETDSIHIEY</i>	<i>JUNB</i>	PXD013649	OMM475	CESC	TSPAPLRTPPRLDGWAHASHWGPGRRRWAPTLGP RGAANHTGLRPSYPAPSPSTSTFTSPPFHFFLYVFFL LET <del>DS</del> IHIEYNIFVLTGRGRGGDRGGAGPAAWYS SPRGHWEGDPRPLPSPLCTV <del>I</del> WKR <del>N</del> THLSKEFILR RVCVCVCLFLLNLFK
<i>SHSKYRTKL</i>	<i>CREG1</i>	PXD004894	MM12	UCEC	<del>RT</del> KL <del>RA</del> DPRQASSIW
<i>YEHRKIQM</i>	<i>EIF3M</i>	PXD013649	OMM475	BLCA	<del>KIQ</del> MHFFLTITFLL
<i>KNLGLRRSI</i>	<i>OR4K1</i>	PXD004894 (1), PXD013649(1)	MM5, OMM475	SKCM	<del>GSL</del> RRSIILN
<i>HLLVARAL</i>	<i>OVOL1</i>	PXD004894	MM12	KIRP	<del>VAR</del> ALGVLL <del>EA</del> PRASRIASQLPGQPTLLQPLTRTPVI RTGAPVPWSP <del>PP</del> WAHVLTQAQQ
<i>TGKKKLNL</i>	<i>PPIP5K2</i>	PXD004894	MM39	UCEC	NLSRSWNFLYL

## Discussion

Studies on mutated genes in cancer have mostly focused on missense mutations; in contrast, stop-loss mutations have been barely investigated. Here we have combined mutation data from different cancer patient cohorts to build a large catalogue of stop-loss mutations, comprising 2,066 cases in 32 different cancer types. This includes data from TCGA and other independent cancer patient cohorts ( $N = 12,224$  patients). Overall, stop-loss mutations are found in approximately one out of ten samples. By aggregating data from thousands of patients we have obtained an unbiased view of their distribution in different types of cancer genes and their level of recurrence. In contrast to another previous catalogue of stop-loss mutations (NonStopDB, Dhamija et al. (2020)), which was enriched in the digestive system and respiratory system, cancer with the most mutations here was uterine corpus endometrial carcinoma. While the NonstopDB is built on mutations from the COSMIC catalogue, our database includes eleven individual studies as well as the full set of mutations called in TCGA data. We did not find stop-loss mutations in the four most recurrently mutated genes in NonStopDB (*ACO2*, *PRKCH*, *SMAD4*, and *CDKN2A*), and *PTMA* was only mutated twice in NonStopDB. These strong differences reflect the distinct sample composition of the two catalogues.

In our study, almost 10% of the mutations were recurrent, suggesting that some of them have oncogenic effects. Mutation recurrence is a hallmark of cancer-driver genes (Tamborero et al., 2013). Oncogenes, which promote tumor growth, arise due to increases in transcript abundance or by specific mutations that make them hyperactive. For example, 90% of the B-raf proto-oncogene (*BRAF*) mutations are a substitution of a V by an E in position 600 and cause continuous cell proliferation (Díaz-Jullien et al., 1996). Protein-inactivating mutations are also very frequent in tumor suppressor genes. A recent study testing the effect of stop-loss mutations in protein levels suggests that this kind of mutation can generate hydrophobic degron motifs that promote increased degradation of the protein (Ghosh et al., 2024). In



*SMAD4*, stop-loss mutations generating such a sequence induce the loss of the protein (Dhamija et al., 2020)), similar to the tumor suppressor gene *VHL* (Dhamija et al., 2020; Pal et al., 2025)). Hydrophobic extensions are also more likely to be loaded onto the MHC-I complex by proteasomal degradation (Seong and Matzinger, 2004). Another recent study related hydrophobic extensions from stop-loss to protein loss due to destabilization of the protein (Ghosh et al., 2024). In line with this, we found that extensions from stop-loss mutations are more hydrophobic than the canonical sequences.

We have discovered a highly recurrently mutated gene, *PTMA*, with a stop-loss mutation found in 14 different patient samples. For comparison, the gene *SMAD4* was found mutated in 9 different patients in NonStopDB. Interestingly, whereas the gene *SMAD4* harbored six different stop-loss extension mutations, *PTMA* had the same mutation in all the samples. This suggests that the glutamine that is being created as a direct result of this mutation is important for the function of the extended sequence. The protein extension, which spans 9 amino acids, has several highly charged amino acids (3 lysines and 1 glutamic acid), which is reminiscent of the composition of the wild-type C-terminus, including a previously defined nuclear localization signal (Manrow et al., 1991).

*PTMA* encodes a highly acidic protein, prothymosin alpha (proT $\alpha$ ), which has proliferative activity. The protein migrates to the nucleus where it binds to histones and cooperates with nucleosome assembly (Díaz-Jullien et al., 1996). High expression in different cancer types is associated with oncogenic activity. Extracellularly it enhances the immune system by promoting dendritic cell maturation, binding to Toll-like receptor 4 on macrophages and neutrophils (Romani et al., 2004; Samara et al., 2013), and increasing T lymphocyte activity in combination with interleukin-2 (Voutsas et al., 2000). In ovarian tumors, proT $\alpha$  inhibits growth by restoring immunosuppressed lymphocytes (Voutsas et al., 2013). ProT $\alpha$  is further the precursor of Thymosin alpha 1 by cleavage at the N-terminal by asparaginyl endopeptidase

(Chen et al., 2006). Similar to proT $\alpha$ , thymosin alpha 1 has been associated with immune regulation by enhancing T cell differentiation, anti-body, chemokine production (Li et al., 2010) and decreasing tumor growth in different animal models (Beuth et al., 2000; Chen et al., 2006; Moody, 2007; Qin et al., 2009; Sungarian et al., 2009).

Analysis of transcriptomics data from TCGA showed that *PTMA* was overexpressed in several cancer types, with strong activation in testicular germ cell tumors, where half the stop-loss cases had been identified. We did not observe any significant differences in the expression of the *PTMA* transcript in samples carrying the stop-loss mutations versus the rest of the samples. Taken together, these results suggest that the mutation increases the protein's activity which, added to the increased expression, contributes to oncogenicity.

Previously, we reported an association of stop-loss mutations with immune checkpoint inhibitor response in bladder cancer patients (Boll et al., 2025). An explanation for this observation is the potential generation of tumor-specific neoantigens from the peptide extensions derived from stop-loss. We searched two immunopeptidomics datasets to evaluate if tumor cells present these peptide extensions. We found several candidates of HLA-I-bound peptides that correspond to protein extensions predicted for stop-loss mutations in cancer samples. While still preliminary, these results support our hypothesis that stop-loss extensions can be a previously neglected source of tumor-specific antigens with potential implications in immunotherapy and personalized cancer vaccine development.

Our results highlight the potential role of stop-loss mutations in cancer. Although comparatively infrequent, we found several recurrently mutated genes over different cancer types. Future studies could focus on validating the immunogenic potential of stop-loss-derived peptides through experimental approaches combining immunopeptidomics with functional T cell activation assays. Additionally, investigating the prevalence and impact of these neoantigens across different cancer types could provide broader insights into their role in immune recognition.

## Methods

### Somatic mutation datasets

For this study, we have processed raw whole exome sequencing data from paired germline and tumor samples of six datasets (IMvigor210, HdM-BLCA-1, MIAO-2018, SNY-2017, Chong, Kraemer). Sequencing reads were trimmed with cutadapt (v4.1) and quality-checked using FastQC (version 0.11.7). Reads were aligned to GRCh38 using BWA (version 0.7.17) and, in cases of several fastq files per patient, merged with GATK MergeSamFiles. Base recalibration and duplicate marking were done with GATK BaseRecalibrator. Contamination was estimated with GATK CalculateContamination, and coverage was assessed using Qualimap (version 2.2.1). Alignment metrics were collected with GATK CollectAlignmentSummaryMetrics. Mutations were called using GATK Mutect2, Strelka2 (v2.9.10), and VarScan2 (version 2.4.4), with SAMtools (version 1.12) mpileup providing input for VarScan2. Mutect2 used the germline-resource file somatic-hg38\_af-only492 gnomad.hg38.vcf.gz, and filtering was done with GATK FilterMutectCalls. Only PASS mutations were retained. An ensemble mutation file, requiring detection by at least two callers, was generated using bcbio-variation-ensemble (version 0.2.2.6) and annotated with VEP (version 104). We generated maf files using vcf2maf (version 0.1.16). Finally, mutations were filtered for a population-wide allele frequency of  $< 5\%$  (gnomAD), sample depth  $\geq 30X$ , and alternative allele depth  $\geq 3X$ . Called mutation tables of five additional datasets were downloaded (Bassani-Sternberg (Bassani-Sternberg et al., 2016), Loeffler (Löffler et al., 2019), Roper/Yue (Roper et al., 2020), FIS\_20200101 (medRxiv 2024.06.28.24309634), one of which only included mutations from a mutation panel of 468 cancer-related genes (Samstein et al., 2019). Finally, we have downloaded all available somatic mutation data of the 31 TCGA cohorts. An overview of all datasets is presented in Table 3.1, with Table S1 containing more detailed information on the TCGA-derived data.

The gene sets of tested tumor suppressors and oncogenes were obtained from COSMIC (<https://cancer.sanger.ac.uk/cosmic/census?tier=2>, last accessed September 11th, 2023). Other cancer-related genes used for the third gene list include DNA-damage repair, chromatin-modifying, and chromatin-regulating genes (Nordentoft et al., 2014; Robertson et al., 2017).

### Mutation clonality

Processing the missense mutation of all patients with at least one stop-loss mutation, we first, calculated the cancer cell fraction (CCF) as  $CCF = VAF / p * (2 * (1 - p) + c * p)$  with  $c$  being the copy number at the mutation position and  $p$  the tumor sample purity. For a majority of stop-loss and missense mutations, no copy number and/or tumor sample purity information was available. We therefore used the subset of mutations for which we obtained a CCF, to classify mutations with a value above 0.9 as clonal. Then, we used this classification to obtain the optimal tumor variant allele frequency threshold that separates clonal and subclonal within this subset and applied this tumor variant allele frequency threshold to all missense and stop-loss mutations. The optimal thresholds used were  $t\_vaf=0.22$  for the stop-loss mutations and  $t\_vaf=0.21$  for the rest of the missense.

### Identification of tumor-specific antigens derived from stop-loss mutations

We generated a database containing all possible extensions of canonical proteins generated by stop-loss mutations ( $n = 17,356$  proteins). For 1282 transcripts, no downstream stop codon was found (7.39%). Each stop-loss mutation results in a codon that encodes one of six or seven different amino acids, depending on the specific stop codon affected (Figure S1). Therefore, we ended up with 100,035 potential, unique peptide extensions. The code for the extensions prediction has been uploaded to Git Hub (<https://github.com/justalilibit/stop-loss->

muta\_in\_canonical). To validate these peptides derived from stop-loss mutations and their capability to bind HLA receptors, we downloaded publicly available mass spectrometry immunopeptidomics data from 24 patients with metastatic malignant melanoma (PXD004894, Bassani-Sternberg et al. (2016)), as well as 3 melanoma cell lines from patient-derived tumors and 4 primary melanoma cell lines (PXD013649, Chong et al. (2020)). For patient MM16, the pipeline failed.

The search was performed with MHCquant v2.6 (Bichmann et al., 2019), an nf-core pipeline implemented in Nextflow (Ewels et al., 2020) that uses Comet as a search engine and percolator for peptide identification. A customized database was built by combining the annotated human proteome sourced from Swiss-Prot/TrEMBL, including isoforms (comprising 103,789 sequences, downloaded on 21<sup>st</sup> of April 2023) and the 100,035 peptide sequences from the 16,074 genes in which we mutated the stop codon. To reduce redundancies, we removed the canonical transcripts in which the stop-loss mutation was transformed, resulting in a total of 188,500 sequences. A false discovery rate (FDR) of 1% was applied.

### Hydrophobicity and isoelectric point

We used the R package ‘Peptides’ (version 2.4.6) to calculate the GRAVY hydrophobicity index with the “KyteDoolittle” scale, and the isoelectric point with the EMBOSS pK-scale.

### Predicting intronic ORFs

We downloaded the database of canonical proteins from UniProt and SwissProt to compare the properties of peptide extensions from stop-loss mutations with canonical sequences and intronic regions. Further, we predicted iORFs in human introns (Genome assembly hg38) using orfipy (version 0.0.4), using a minimum peptide length of 9 amino acids.

## Gene expression data

Gene expression values were directly downloaded from the TCGA and GTEx projects. Whenever possible, we used matched cancer/control gene expression data. For testicular cancer (TCGT), TCGA does not contain adjacent data. Instead, we used healthy testis expression data from the GTEx project.

## Statistical analysis

Plots were generated using Python (version 3.8.6), R (version 4.1.2), and Rstudio (version 1.4). When testing for enrichment between two groups the Chi Square test was performed.

## References

- Arribere, J. A., Cenik, E. S., Jain, N., Hess, G. T., Lee, C. H., Bassik, M. C., and Fire, A. Z. (2016). Translation readthrough mitigation. *Nature*, 534(7609):719–723. Publisher: Nature Publishing Group.
- Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., H. Busch, D., Peschel, C., Rad, R., Cox, J., Mann, M., and Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications*, 7(1):13404. Publisher: Nature Publishing Group.
- Beuth, J., Schierholz, J. M., and Mayer, G. (2000). Thymosin 11 application augments immune response and down-regulates tumor weight and organ colonization in BALB/c-mice. *Cancer Letters*, 159(1):9–13.
- Bichmann, L., Nelde, A., Ghosh, M., Heumos, L., Mohr, C., Peltzer, A., Kuchenbecker, L., Sachsenberg, T., Walz, J. S., Stevanović, S., Ramnensee, H.-G., and Kohlbacher, O. (2019). MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics. *Journal of Proteome Research*, 18(11):3876–3884.
- Boll, L. M., Vázquez Montes de Oca, S., Camarena, M. E., Castelo, R., Bellmunt, J., Perera-Bel, J., and Albà, M. M. (2025). Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts. *Nature Communications*, 16(1):1213. Publisher: Nature Publishing Group.
- Chen, P. F., Fu, G. F., Zhang, H. Y., Xu, G. X., and Hou, Y. Y. (2006). Liposomal plasmid DNA encoding human thymosin 1 and interferon 1 potently inhibits liver tumor growth in ICR mice.

*Journal of Gastroenterology and Hepatology*, 21(10):1538–1543.  
\_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1746.2006.04536.x>.

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G., and Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*, 11(1):1293.

Dhamija, S., Yang, C. M., Seiler, J., Myacheva, K., Caudron-Herger, M., Wieland, A., Abdelkarim, M., Sharma, Y., Riester, M., Groß, M., Maurer, J., and Diederichs, S. (2020). A pan-cancer analysis reveals nonstop extension mutations causing SMAD4 tumour suppressor degradation. *Nature Cell Biology*, 22(8):999–1010. Publisher: Nature Publishing Group.

Díaz-Jullien, C., Pérez-Estévez, A., Covelo, G., and Freire, M. (1996). Prothymosin alpha binds histones in vitro and shows activity in nucleosome assembly assay. *Biochimica Et Biophysica Acta*, 1296(2):219–227.

Enkemann, S., Wang, R.-H., Trumbore, M., and Berger, S. (2000). Functional discontinuities in prothymosin caused by caspase cleavage apoptotic cells. *Journal of Cellular Physiology*, 182(2):256–268.

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, 38(3):276–278. Publisher: Nature Publishing Group.

Ferreira, A. M., Tuominen, I., van Dijk-Bos, K., Sanjabi, B., van der Sluis, T., van der Zee, A. G., Hollema, H., Zazula, M., Sijmons, R. H., Aaltonen, L. A., Westers, H., and Hofstra, R. M. W. (2014). High



frequency of RPL22 mutations in microsatellite-unstable colorectal and endometrial tumors. *Human Mutation*, 35(12):1442–1445.

Flores, S., Prego, M., Smart, T., Wang, J., Agapite, J., Arun, S., Chowning, M., Gramatikova, S., Hohlt, K., Li, L., Ross, R., Arya, P., and Sanapareddy, N. (2024). P590: Evaluation of stop-loss and frameshift variants extending translation beyond the reference stop codon detected by carrier screening. *Genetics in Medicine Open*, 2:101496.

García-Aznar, J. M., Alonso Alvarez, S., and Bernal del Castillo, T. (2024). Pivotal role of BCL11B in the immune, hematopoietic and nervous systems: a review of the BCL11B-associated phenotypes from the genetic perspective. *Genes & Immunity*, 25(3):232–241. Publisher: Nature Publishing Group.

Ghosh, A., Riester, M., Pal, J., Lainde, K.-A., Tangermann, C., Wanning, A., Dueren, U. K., Dhamija, S., and Diederichs, S. (2024). Suppressive cancer nonstop extension mutations increase C-terminal hydrophobicity and disrupt evolutionarily conserved amino acid patterns. *Nature Communications*, 15(1):9209. Publisher: Nature Publishing Group.

Goodenough, E., Robinson, T. M., Zook, M. B., Flanigan, K. M., Atkins, J. F., Howard, M. T., and Eisenlohr, L. C. (2014). Cryptic MHC class I-binding peptides are revealed by aminoglycoside-induced stop codon read-through into the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15):5670–5675.

Jiang, X., Kim, H.-E., Shu, H., Zhao, Y., Zhang, H., Kofron, J., Donnelly, J., Burns, D., Ng, S.-c., Rosenberg, S., and Wang, X. (2003). Distinctive Roles of PHAP Proteins and Prothymosin- in a Death Regulatory Pathway. *Science*, 299(5604):223–226. Publisher: American Association for the Advancement of Science.

Klauer, A. A. and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley interdisciplinary*

reviews. *RNA*, 3(5):649–660.

- Li, J., Liu, C. H., and Wang, F. S. (2010). Thymosin alpha 1: Biological activities, applications and genetic engineering production. *Peptides*, 31(11):2151–2158.
- Löffler, M. W., Mohr, C., Bichmann, L., Freudenmann, L. K., Walzer, M., Schroeder, C. M., Trautwein, N., Hilke, F. J., Zinser, R. S., Mühlenthalbruch, L., Kowalewski, D. J., Schuster, H., Sturm, M., Matthes, J., Riess, O., Czernmel, S., Nahnsen, S., Königsrainer, I., Thiel, K., Nadalin, S., Beckert, S., Bösmüller, H., Fend, F., Velic, A., Maček, B., Haen, S. P., Buonaguro, L., Kohlbacher, O., Stevanović, S., Königsrainer, A., and Rammensee, H.-G. (2019). Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Medicine*, 11(1):28.
- Malicet, C., Giroux, V., Vasseur, S., Dagorn, J. C., Neira, J. L., and Iovanna, J. L. (2006). Regulation of apoptosis by the p8/prothymosin complex. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8):2671–2676.
- Manrow, R. E., Sburlati, A. R., Hanover, J. A., and Berger, S. L. (1991). Nuclear targeting of prothymosin alpha. *Journal of Biological Chemistry*, 266(6):3916–3924.
- Moody, T. W. (2007). Thymosin 1 as a Chemopreventive Agent in Lung and Breast Cancer. *Annals of the New York Academy of Sciences*, 1112(1):297–304. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1415.040>.
- Munro, E. G., Jain, M., Oliva, E., Kamal, N., Lele, S. M., Lynch, M. P., Guo, L., Fu, K., Sharma, P., Remmenga, S., Growdon, W. B., Davis, J. S., Rueda, B. R., and Batra, S. K. (2009). Upregulation of MUC4 in cervical squamous cell carcinoma: pathologic significance. *International Journal of Gynecological Pathology: Official Journal of the International Society of Gynecological Pathologists*, 28(2):127–133.

Nordentoft, I., Lamy, P., Birkenkamp-Demtröder, K., Shumansky, K., Vang, S., Hornshøj, H., Juul, M., Villesen, P., Hedegaard, J., Roth, A., Thorsen, K., Høyer, S., Borre, M., Reinert, T., Frstrup, N., Dyrskjød, L., Shah, S., Pedersen, J. S., and Ørntoft, T. F. (2014). Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer. *Cell Reports*, 7(5):1649–1663.

Novetsky, A. P., Zighelboim, I., Thompson, D. M., Powell, M. A., Mutch, D. G., and Goodfellow, P. J. (2013). Frequent Mutations in the RPL22 Gene and its Clinical and Functional Implications. *Gynecologic oncology*, 128(3):10.1016/j.ygyno.2012.10.026.

Pal, J., Riester, M., Ganner, A., Ghosh, A., Dhamija, S., Mookherjee, D., Voss, C., Frew, I. J., Kotsis, F., Neumann-Haefelin, E., Spang, A., and Diederichs, S. (2025). Nonstop mutations cause loss of renal tumor suppressor proteins VHL and BAP1 and affect multiple stages of protein translation. *Science Advances*, 11(7):eadr6375. Publisher: American Association for the Advancement of Science.

Pérez-Estévez, A., Díaz-Jullien, C., Covelo, G., Salgueiro, M. T., and Freire, M. (1997). A 180-kDa protein kinase seems to be responsible for the phosphorylation of prothymosin alpha observed in proliferating cells. *The Journal of Biological Chemistry*, 272(16):10506–10513.

Qin, Y., Chen, F.-D., Zhou, L., Gong, X.-G., and Han, Q.-F. (2009). Proliferative and anti-proliferative effects of thymosin 1 on cells are associated with manipulation of cellular ROS levels. *Chemico-Biological Interactions*, 180(3):383–388.

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., Hinoue, T., Laird, P. W., Hoadley, K. A., Akbani, R., Castro, M. A. A., Gibb, E. A., Kanchi, R. S., Gordenin, D. A., Shukla, S. A., Sanchez-Vega, F., Hansel, D. E., Czerniak, B. A., Reuter, V. E., Su, X., de Sa Carvalho, B., Chagas, V. S., Mungall, K. L., Sadeghi, S., Pedamallu, C. S., Lu, Y., Klimczak, L. J., Zhang, J., Choo, C., Ojesina, A. I., Bullman, S., Leraas, K. M., Lichtenberg, T. M., Wu,

- C. J., Schultz, N., Getz, G., Meyerson, M., Mills, G. B., McConkey, D. J., TCGA Research Network, Weinstein, J. N., Kwiatkowski, D. J., and Lerner, S. P. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell*, 171(3):540–556.e25.
- Romani, L., Bistoni, F., Gaziano, R., Bozza, S., Montagnoli, C., Perruccio, K., Pitzurra, L., Bellocchio, S., Velardi, A., Rasi, G., Di Francesco, P., and Garaci, E. (2004). Thymosin alpha 1 activates dendritic cells for antifungal Th1 resistance through toll-like receptor signaling. *Blood*, 103(11):4232–4239.
- Roper, N., Brown, A.-L., Wei, J. S., Pack, S., Trindade, C., Kim, C., Restifo, O., Gao, S., Sindiri, S., Mehrabadi, F., El Meskini, R., Ohler, Z. W., Maity, T. K., Venugopalan, A., Cultraro, C. M., Akoth, E., Padiernos, E., Chen, H., Kesarwala, A., Smart, D. K., Nilubol, N., Rajan, A., Piotrowska, Z., Xi, L., Raffeld, M., Panchenko, A. R., Sahinalp, C., Hewitt, S., Hoang, C. D., Khan, J., and Guha, U. (2020). Clonal Evolution and Heterogeneity of Osimertinib Acquired Resistance Mechanisms in EGFR Mutant Lung Cancer. *Cell Reports Medicine*, 1(1):100007.
- Samara, P., Ioannou, K., Neagu, M., Arnogiannaki, N., Ardavanis, A., Voelter, W., and Tsitsilonis, O. (2013). The C-terminal decapeptide of prothymosin is responsible for its stimulatory effect on the functions of human neutrophils in vitro. *International Immunopharmacology*, 15(1):50–57.
- Samara, P., Ioannou, K., and Tsitsilonis, O. E. (2016). Chapter Eight - Prothymosin Alpha and Immune Responses: Are We Close to Potential Clinical Applications? In Litwack, G., editor, *Vitamins and Hormones*, volume 102 of *Thymosins*, pages 179–207. Academic Press.
- Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., Barron, D. A., Zehir, A., Jordan, E. J., Omuro, A., Kaley, T. J., Kendall, S. M., Motzer, R. J., Hakimi, A. A., Voss,

- M. H., Russo, P., Rosenberg, J., Iyer, G., Bochner, B. H., Bajorin, D. F., Al-Ahmadie, H. A., Chaft, J. E., Rudin, C. M., Riely, G. J., Baxi, S., Ho, A. L., Wong, R. J., Pfister, D. G., Wolchok, J. D., Barker, C. A., Gutin, P. H., Brennan, C. W., Tabar, V., Mellingerhoff, I. K., DeAngelis, L. M., Ariyan, C. E., Lee, N., Tap, W. D., Gounder, M. M., D'Angelo, S. P., Saltz, L., Stadler, Z. K., Scher, H. I., Baselga, J., Razavi, P., Klebanoff, C. A., Yaeger, R., Segal, N. H., Ku, G. Y., DeMatteo, R. P., Ladanyi, M., Rizvi, N. A., Berger, M. F., Riaz, N., Solit, D. B., Chan, T. A., and Morris, L. G. T. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics*, 51(2):202–206. Publisher: Nature Publishing Group.
- Scholz, S. L., Horn, S., Murali, R., Möller, I., Sucker, A., Sondermann, W., Stiller, M., Schilling, B., Livingstone, E., Zimmer, L., Reis, H., Metz, C. H., Zeschnigk, M., Paschen, A., Steuhl, K.-P., Schadendorf, D., Westekemper, H., and Griewank, K. G. (2015). Analysis of SDHD promoter mutations in various types of melanoma. *Oncotarget*, 6(28):25868–25882.
- Segade, F. and Gómez-Márquez, J. (1999). Prothymosin alpha. *The International Journal of Biochemistry & Cell Biology*, 31(11):1243–1248.
- Seong, S.-Y. and Matzinger, P. (2004). Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nature Reviews Immunology*, 4(6):469–478. Publisher: Nature Publishing Group.
- Shibata, N., Ohoka, N., Sugaki, Y., Onodera, C., Inoue, M., Sakuraba, Y., Takakura, D., Hashii, N., Kawasaki, N., Gondo, Y., and Naito, M. (2015). Degradation of Stop Codon Read-through Mutant Proteins via the Ubiquitin-Proteasome System Causes Hereditary Disorders. *The Journal of Biological Chemistry*, 290(47):28428–28437.
- Sungarian, A., Cielo, D., Sampath, P., Bowling, N., Moskal, P., Wands, J. R., and de la Monte, S. M. (2009). Po-

tential Role of Thymosin-1 Adjuvant Therapy for Glioblastoma. *Journal of Oncology*, 2009(1):302084. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2009/302084>.

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M. S., Getz, G., Bader, G. D., Ding, L., and Lopez-Bigas, N. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Scientific Reports*, 3(1):2650. Publisher: Nature Publishing Group.

Trumbore, M. W., Wang, R.-H., Enkemann, S. A., and Berger, S. L. (1997). Prothymosin *in Vivo* Contains Phosphorylated Glutamic Acid Residues\*. *Journal of Biological Chemistry*, 272(42):26394–26404.

Tsitsiloni, O. E., Heimer, E., Felix, A., Yialouris, P. P., Vamvoukakis, J., Voelter, W., and Haritos, A. A. (1994). Radioimmunoassays for the C-terminus of prothymosin and the N-terminus of parathymosin for the measurement of the levels of -thymosins in human cancer. *Journal of Immunological Methods*, 169(2):163–171.

Voutsas, I., Baxevas, C., Gritzapis, A., Missitzis, I., Stathopoulos, G., Archodakis, G., Banis, C., Voelter, W., and Papamichail, M. (2000). Synergy between interleukin-2 and prothymosin for the increased generation of cytotoxic T lymphocytes against autologous human carcinomas. *Cancer Immunology Immunotherapy*, 49(8):449–458.

Voutsas, I., Pistamaltzian, N., Tsiatas, M., Skopeliti, M., Katsila, T., Mavrothalassiti, I., Spyrou, S., Dimopoulos, M.-A., Tsitsilonis, O., and Bamias, A. (2013). Ovarian malignant ascites-derived lymphocytes stimulated with prothymosin or its immunoactive decapeptide lyse autologous tumour cells in vitro and retard tumour growth in SCID mice. *European Journal of Cancer*, 49(7):1706–1714.

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., and Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature Genetics*, 46(11):1160–1165. Publisher: Nature Publishing

Group.

Weinstein, H. N. W., Hu, K., Fish, L., Chen, Y.-A., Allegakoen, P., Pham, J. H., Hui, K. S. F., Chang, C.-H., Tutar, M., Benitez-Rivera, L., Baco, M. B., Song, H., Giacomelli, A. O., Vazquez, F., Ghandi, M., Goodarzi, H., and Huang, F. W. (2024). RPL22 is a tumor suppressor in MSI-high cancers and a splicing regulator of MDM4. *Cell Reports*, 43(8):114622.

Xia, P., Choi, A. H., Deng, Z., Yang, Y., Zhao, J., Wang, Y., Hardwidge, P. R., and Zhu, G. (2016). Cell membrane-anchored MUC4 promotes tumorigenicity in epithelial carcinomas. *Oncotarget*, 8(8):14147–14157. Publisher: Impact Journals.





Cancer remains one of the leading global health burdens, characterized by its immense heterogeneity across and within tumor types. This complexity necessitates tailored treatment approaches to effectively address the diverse biological mechanisms driving cancer. Among emerging therapies, ICIs as part of immunotherapy have revolutionized cancer treatment. ICIs work by enhancing the ability of the immune system to recognize and eliminate cancer cells. Under normal conditions, immune checkpoints, such as PD-1/PD-L1 and CTLA-4, act as regulatory mechanisms to prevent overactivation of the immune system and avoid damage to healthy tissues. However, many cancers exploit these pathways by overexpressing checkpoint proteins, effectively suppressing T cell activity and evading immune detection. ICI therapies, such as anti-PD-1, anti-PD-L1, and anti-CTLA-4 antibodies, block these inhibitory signals, reactivating T cells and their ability to target tumor cells.

ICIs offer remarkable benefits for some patients. However, despite their success, predicting which patients will respond to ICIs remains a significant challenge. Our understanding of the underlying drivers of immunotherapy response is still limited, underscoring the urgent need for deeper insights into tumor-immune interactions and reliable biomarkers to guide treatment decisions.

#### **4.1 From small to large-scale analyses of ICI response**

The first two parts of this thesis focus on the biomarker discovery in urothelial bladder cancer in response to the ICI anti-PD-L1/anti-PD-1. While the first chapter handles a small dataset of less than 30 patients

(referred to as HdM-BLCA-1), we were still able to obtain meaningful insights from the analysis.

First, we confirmed that a high number of somatic mutations is associated with better outcomes of ICI treatment. A high TMB increases the likelihood of producing tumor-specific neoantigens, which can be recognized by T cells, triggering an immune response.

Tumors with high TMB often exhibit greater responsiveness to ICI, as they present more targets for abnormal antigen presentation initiating immune recognition. In the first dataset, we analyzed, clonal TMB better separated the two response groups than total TMB, suggesting that cancer heterogeneity negatively relates to ICI response. This is coherent with studies in melanoma and non-small cell lung cancer (McGranahan et al., 2016; Wolf et al., 2019), as well as a dataset of urothelial bladder cancer by Miao *et al.* (referred to as MIAO-2018) (Miao et al., 2018). We extended our analysis by including the bladder cancer cohort SNYDER-2017 (Snyder et al., 2017). Although clonality could not be calculated, we showed that high variant allele frequency significantly separates treatment responders and non-responders. Compared with this, the meta-analysis of the six different cohorts also showed more clonal non-synonymous mutations among the response group, but we did not find clonal TMB to be a better predictor for response than the total TMB. This meta-analysis also includes the MIAO-2018 dataset, where we were unable to replicate their result of responders having a significantly higher clonal TMB. Similar results were obtained for the predicted neoantigen burden. While the number of neoantigens originating from clonal mutations showed a stronger separation of responders and non-responders in the smaller cohort, this was not found in the combined cohort. This highlights how small datasets can include subgroups where specific effects appear significant but are no longer observed in larger cohorts. Moreover, differences in computational pipelines regarding data processing, variant calling, filtering, and quality control may strongly influence results as shown by the different results in the MIAO-2018

dataset. Particularly in smaller datasets such variations in one or two patients can markedly impact significance.

Besides, the computational prediction of immunogenic antigens faces several challenges. Several studies report that a small proportion of their *in situ* predicted neoantigens were found to be immunogenic *in vitro* (Bassani-Sternberg et al., 2016; Yadav et al., 2014; Zhang et al., 2017). This divergence is impacted by biological processes that are difficult to forecast with computational pipelines. Today's prediction programs focus mostly on peptide-MHC binding, while recognition of the TCR is crucial to elicit an immune response (Müller et al., 2017). The TCR-peptide binding however is little understood and therefore difficult to predict. Furthermore, the presentation of antigens includes additional downstream peptide processing steps, including proteolysis, the localization of the peptide in the endoplasmic reticulum, and the loading onto the MHC.

Mutation instability in bladder cancer is often connected to the APOBEC mutational signature. While in the HdM-BLCA-1 cohort, we only saw a significant association between APOBEC and ICI response in clonal mutations, combining several datasets showed that responders are enriched in APOBEC-induced mutations. Again, an increase in sample size strongly affected statistical significance.

The power of a larger sample size is further visible when analyzing the composition of the TME and its association with treatment response. While previous studies have repeatedly outlined the importance of T cell infiltration as a key biomarker for ICI response (Hammerl et al., 2021; Hernando-Calvo et al., 2023; Litchfield et al., 2020; Mariathasan et al., 2018), we were only able to report tendencies in the HdM-BLCA-1 cohort. In contrast, the larger number of RNA-Seq samples obtained later allowed us to get a clearer picture of the key players. The abundance of immune cells such as CD8<sup>+</sup> T cells and M1 macrophages, as well as the expression of IFN- $\gamma$  and its related gene signatures, were found to be strong predictors of ICI response in bladder cancer. In comparison, high TGF- $\beta$  values reduce treatment response as also

reported by others. This underlies its anti-inflammatory effect, limiting T cell proliferation while inducing EMT and Treg population. Similarly, markers for EMT and stroma cells, indicating angiogenesis, tumor migration, and T cell exclusion, negatively relate to ICI response.

So, in general, tumors with high levels of immune cell infiltration, particularly of CD8+ T cells, are better positioned for immune-mediated cancer cell detection. Interestingly, patients of the two subtypes with the highest immune infiltration, basal-squamous and luminal-infiltrated, do not respond better than the three non-immune-infiltrated subtypes. Within the group of high CD8+ infiltrated patients, we found that the immune suppressive TGF- $\beta$  values separate the patients into less and more likely to respond to treatment. We hypothesize that immune-infiltrate tumors do not show higher response rates than non-immune-infiltrated tumors due to the countering of immune-suppressive pathways, such as TGF- $\beta$ , that hinder a successful immune response. The non-immune-infiltrated subtypes showed not just higher levels for markers of immune suppression, but also an overall lower TMB, which negatively affects treatment response. The only subtype with a significantly better response to ICI was the rare neuronal subtype, potentially due to its particularly low levels of TGF- $\beta$ . Interestingly, the antigen-presenting machinery was lowly expressed in the six non-responders of this group, suggesting that their tumors show a decreased ability to present neoantigens, a critical step for ICI efficacy.

## 4.2 Biomarkers driving response differ between subgroups

Not all patients benefit from ICIs, highlighting the need to identify robust biomarkers and understand the mechanisms driving treatment response and resistance. And even though scientists have made a great effort to characterize each of these mechanisms and find strong biomarkers, predicting response to immunotherapy is still a riddle to be solved. Recent breakthroughs in omics technologies opened new

opportunities for biomedical research. Advancements in genomics, transcriptomics, proteomics, and metabolomics facilitate the exploration of biological systems with unparalleled precision and depth. The growth in high-dimensional biological data, coupled with the rapid expansion of computational resources, has paved the way for the development and application of computational models, offering powerful tools to interpret and unravel the complexities of biological systems.

In the cohort presented in the first publication of this thesis, we found clonal TMB to be a better separator of response groups than general non-synonymous TMB. This was not the case when several cohorts were combined in the second results section. Examples like this can be found all over the literature where observations hold true in one but not another cohort. The heterogeneity of tumor patients asks for a multi-factorial approach. Previous work has suggested that immune-infiltrated patients not only respond better but can also balance out a low TMB. We have seen that subtypes with high immune infiltration do not respond better by default, hypothesizing that this might be due to the immune-suppressive effects of pathways such as TGF- $\beta$ . In this work, we find that different molecular subtypes of advanced bladder cancer have distinct predictive markers for response that represent a combination of known biomarkers.

We selected the features for the prediction model based on the previous analysis. These covered mutation-based variables (TMB, non-stop mutations, APOBEC-enrichment score), a signature of tumor-specific lncRNAs, gene expression of selected genes (*PD-1*, *PD-L1*, *CCND1*), immune signatures (IFN- $\gamma$ , stroma/EMT, inflamed T cells, TGF- $\beta$  and antigen-presenting machinery pathways), immune cell abundance (M1 macrophages, CD4 memory activated T cells, CD8 T cells and regulatory T cells), and clinical information (ECOG, liver metastasis). The random forest model built with the full set of features reached an Area under the curve (AUC) of 0.761. In comparison, a random forest model using only TMB was much less accurate in response

prediction (AUC=0.678). Applying the 10 mutations/Mb threshold approved by the FDA for solid tumors, achieved an even lower accuracy (AUC=0.61), showing that considering markers of different biological processes adds to the prediction of ICI response. For a portion of the patients, only a TMB estimate and no further mutation-based data was available. In a clinical setting, running WES and RNA sequencing is cost- and time-intensive. We have therefore built a model using only the features based on RNA-Seq data together with the TMB, reaching a similar AUC as when using all features (AUC=0.747), but being more robust as it was built on a bigger sample size. This is also reflected in the high response prediction accuracy when validated on an external cohort (AUC=0.764).

Remarkably, a model built using only high immune-infiltrated tumor patients (Basal-squamous and luminal-infiltrated subtypes) showed the highest accuracy (AUC=0.761), while the selected features showed only a modest ability to predict ICI response in the non-immune-infiltrated group. This lets us to conclude that drivers of ICI treatment response in immune-infiltrated tumors are much better understood than those of the other subtypes. We have therefore focussed on the non-immune-infiltrated subtypes separately and built a random forest model on the biggest subtype, luminal-papillary, and decision trees for the two smaller subtypes luminal and neuronal. In the luminal subtype, we found that besides TMB and M1 macrophage infiltration, either high expression of PD-1 or low stroma/EMT values were characteristic for treatment responders. In the luminal subtype, we found most non-responders to have high infiltration levels of regulatory T cells. Finally, in the decision tree of patients with the neuronal subtype, the most important feature separating response was the expression of antigen-presenting machinery genes, as discussed above.

Future research and clinical study designs should consider the different subtypes and their immune-infiltration status and prioritize unraveling the mechanisms driving response in non-immune-infiltrated tumors. A better understanding of the differences driving response in this group

of patients could be key to improving patient health, for example by combining immunotherapy with therapeutic interventions increasing local immune invasion. Potential drug candidates could be CD8+ T cell recruiting cytokine signals, such as seasonal flu vaccines (Newman et al., 2020), the CDK4/6 inhibitor abemaciclib (Zhang et al., 2020), or COX inhibitors (Bronger et al., 2012).

### 4.3 Stop codons impact cancer

In recent years, the outlook for novel sources of tumor-specific neoantigens has become a major focus in cancer research, aiming to develop targeted therapies in immunotherapy and cancer vaccines. While missense mutations have been the most extensively studied, frameshift mutations have also been proposed as contributors to immunogenic neoantigens (Linnebacher et al., 2001; Turajlic et al., 2017). Additionally, recent efforts are directed toward exploring the ‘dark’ genome, investigating elements such as non-coding regions, transposable elements, or unannotated ORFs to uncover additional targets (Camarena et al., 2024). A mutation type that has received little attention is the stop-loss mutation, leading to an extension of the peptide sequence into the non-coding genome. In the second part of this study, we observed that the group of ICI therapy responders showed a significantly larger number of stop-loss mutations than the group of non-responders. We hypothesize that the peptide extensions derived from stop-loss mutations give rise to immunogenic neoantigens in the tumor cells. Compared to other non-synonymous mutations, the effect of stop-loss mutations on the affected gene is little understood. While many tools exist to predict variant effects, the impact of stop-loss mutations is more complex due to the potential effect of the C-terminal extension on the protein’s functionality, highlighting the need for an in-depth characterization of this mutation type.

Stop-loss mutations comprise around 0.2% of mutations (Nehrt et al., 2012). To study their effect on cancer, we combined mutation data from 12,224 patients of 32 different cancer types, generating

a database of over 2000 stop-loss mutations. Most stop-loss mutations were found in cancers of the genitourinary, respiratory, and digestive systems. While little is understood about the effect of stop-loss mutations in cancer, we found almost 10% of the genes to be recurrently mutated, suggesting a biological relevance. We also found that oncogenes showed an enrichment of stop-loss mutations compared to missense mutations. *PTMA* stands out with 14 patients carrying the same T>C mutation, half in testicular cancer. The encoded protein, prothymosin alpha, has been described to have proliferating activities and is strongly expressed in many cancer tissues (Enkemann et al., 2000; Jiang et al., 2003; Malicet et al., 2006). While the mutation did not affect the expression in patients with available expression data, we found a strong overexpression of *PTMA* in tumor tissue. This included testicular cancer, which was recurrently *PTMA* stop-loss mutated.

In general, we found that stop-loss mutations are overrepresented in oncogenes and cancer-related genes over missense mutations, suggesting a positive effect of stop-loss on the activity of the gene. In comparison, a recent study on another database of stop-loss mutations (NonStopDB, Dhamija et al. (2020)) reports that when tested experimentally, around half of the extensions derived from stop-loss mutations reduced the protein abundance (Ghosh et al., 2024). Focussing on three tumor suppressor genes, *SMAD4*, *PTEN*, *VHL*, and *BAP1*, they found that the C-terminal extensions cause protein loss, either by degradation or, in the latter case, by translation inhibition of the mRNA (Dhamija et al., 2020; Pal et al., 2025).

Our database offers a complementary perspective on stop-loss mutations. The NonStopDB was built on the Catalogue of Somatic Mutations in Cancer (COSMIC) data, version 78 (Forbes et al., 2016). Similarly to our database, this incorporates individual studies (60%) and a subset of the International Cancer Genome Consortium (ICGC) and TCGA (40%). This database includes 3,412 unique stop-loss mutations. However, the distribution of stop-loss mutations differs significantly between the two databases. While NonStopDB reports the highest



prevalence in digestive system tumors, our database finds the most frequent mutations in uterine cancer. The variations in mutation occurrence arise from differing cohort compositions. The unbiased nature of the data we analyzed, covering all possible stop-loss mutations in the genome, allowed us to identify an enrichment of this mutation type, both in tumor suppressor genes and oncogenes. In addition, the latter group showed an excess of stop-loss mutations versus missense mutations, suggesting that the first type of mutations might have a role in increasing oncogenicity. We have further computed potential peptide sequence extensions derived from stop-loss mutations. We found the average length of stop-loss extension to be comparatively short with 16 amino acids. The value is lower than the average length published based on a bioinformatics analysis (27.8 amino acids (Shibata *et al.*, 2015)), but very similar to the prediction on the NonStopDB (18 amino acids (Dhamija *et al.*, 2020)). Both we and Dhamija *et al.* predicted extensions based on mutations found in tumors, while Shibata *et al.* used all potential protein extensions. The difference in average length could indicate a specific selection in cancerous tissues. Further analysis comparing the structural and functional impact of these extensions may provide insights into their biological significance and potential role in tumorigenesis.

## 4.4 The limits of current models and future directions

Since their first approval in 2014, there has been a race for publishing the most accurate prediction model for ICI response. Many models have been published, some relying on a single data source, others integrating multiple sources, big pan-cancer studies, and smaller cancer-specific models. A fundamental challenge in predictive modeling remains the right trade-off between simplicity and complexity. The two markers approved by the FDA, high TMB and PD-L1 expression, provide a convenient way to stratify patients but fail to account for the full spectrum of immune interactions that dictate response. Even

measuring a seemingly simple biomarker such as TMB can become complex, as highlighted by the attempts of multi-institutional collaboration to harmonize the variability in how TMB is analyzed and applied in the clinical setting (of Cancer Research, 2025).

A recent pan-cancer model using blood tests from over 9,000 patients shows promising results in predicting over-survival to ICI surpassing results obtained with TMB or PD-L1 staining (Yoo et al., 2025). Other research groups published gene signatures trying to separate patients by response based on the mRNA levels of a set of genes (Benguigui et al., 2024; Hernando-Calvo et al., 2023). More complex, multi-omics approaches incorporating genomic, transcriptomic, and immune profiling data offer a more comprehensive view of tumor-immune dynamics. However, the more complex a model becomes, the less practical it may be for clinical translation. Our work aligns with the latter approach—not as a clinical tool but as a means to deepen our understanding of the intricate mechanisms governing immunotherapy response.

One of the greatest obstacles in prediction is the variability of features among patients but also within the same patient over time. Many published prediction models achieve high accuracy in their training and testing cohorts but fail to present an equally high AUC in independent validation cohorts (Damrauer et al., 2022; Litchfield et al., 2021). This discrepancy can result from overfitting but also be strongly affected by different methods to obtain patient samples. Studies have demonstrated that the same patient can exhibit significant differences in immune composition between primary and metastatic lesions. The immunopeptidome repertoire, for example, varies between tumor regions within the same individual (Kraemer et al., 2023).

Beyond the complexity of biological variables, there is the epistemological question of how we define response. Immunotherapy response is not a static characteristic but a dynamic process. While a universal standard is lacking, the field roughly separates into two approaches. The time-to-event metrics, such as progression-free and overall sur-

vival (Damrauer et al., 2022; Nassar et al., 2020; Powles et al., 2020; Rizvi et al., 2015; Rose et al., 2021; Samstein et al., 2019), and response rate classification based on Response evaluation criteria in solid tumors (RECIST) criteria, patients are separated by the tumor lesion growth into complete responders, partial responders, stable disease, and progressive disease (Benguigui et al., 2024; Eroglu et al., 2018; Hellmann et al., 2018; Mariathasan et al., 2018; Robertson et al., 2023; Snyder et al., 2017; Van Allen et al., 2015; Zhang et al., 2022). Both approaches pose their difficulties. While the discussion around progression-free survival as a surrogate for overall survival and a marker for quality of life is ongoing (Booth and Eisenhauer, 2012), the matter is just as unclear for the RECIST categories, due to the ambiguity of the classification of stable disease cases. Again, different groups took different approaches, in some cases no response includes cases in which the tumor is stable, whereas in others it only refers to cases in which there is growth of tumor lesions. Decisions like this one can strongly impact the results of a study, affecting which features a model will prompt as most associated with response and changing the AUC. In this thesis, we excluded patients with stable disease response status. Even though a lower sample size would affect the predictive power of our analysis, we believed we would obtain more meaningful results by reducing potential noise with a clearer separation of the two response groups.

Finally, despite integrating diverse datasets encompassing clinical, expression, and mutation data, even today's most sophisticated models struggle to predict response to ICIs robustly. This suggests that the existing variables may capture only a fraction of the intricate factors governing immunotherapy response and that others, yet unseen, variables play a crucial role. Research has, for example, shown that the gut microbiome influences immune activation and potentially impacts patient response to ICIs (Chaput et al., 2017; Viaud et al., 2013; Vétizou et al., 2015). Other factors such as epigenetics, metagenomics and more detailed clinical metadata on lifestyle factors may add to build better predictive frameworks. Nevertheless, it is unlikely that a single

model will ever fully predict therapy response. Instead, the future may lie in developing personalized, adaptive models incorporating real-time patient data.

Decoding ICI response is not just a technical challenge, but a reflection of our evolving understanding of tumor-immune dynamics. By achieving the objectives outlined in this thesis, we provide novel insights into the underlying mechanisms of ICI response, particularly in bladder cancer. This also opened new questions, such as the driving factors for response in the non-immune-infiltrated subtypes in bladder cancer or the potential of stop-loss-derived neoantigens as treatment therapies. Each step forward refines the ground of knowledge on the hide-and-seek between tumor and immune system and paves the path for a future where precision medicine can deliver truly personalized therapies.

1. High TMB correlates with better ICI response, as it increases the likelihood of neoantigen presentation. Clonal TMB appears to be a stronger predictor of response in the HdM-BLCA-1 dataset. However, we did not observe this effect in the meta-analysis, highlighting the importance of large, balanced datasets.
2. We report that while immune cell infiltration, particularly of CD8+ T cells, and M1 macrophages, is associated with improved response to ICIs, high immune infiltration does not always predict better outcomes, as immune-suppressive factors like TGF- $\beta$  can counteract the benefits of immune activation.
3. Different bladder cancer subtypes show varying responses to ICIs, with the neuronal subtype displaying the highest response rate, possibly due to low TGF- $\beta$  levels.
4. The use of a large amount of clinical samples allowed us to uncover the main biomarkers in non-immune infiltrated subtypes, such as the antigen presenting machinery in the neuronal subtype.
5. We combined data from over 12,000 patients identifying a significant enrichment of stop-loss mutations in oncogenes and cancer-related genes, suggesting that they can enhance oncogenic activity in certain cases.
6. Our analysis of immunopeptidomics data suggests that stop-loss mutation could be a source of tumor-specific antigens, to be exploited for immunotherapy and cancer-vaccine therapies.



## Bibliography

Abuhelwa, A. Y., Kichenadasse, G., McKinnon, R. A., Rowland, A., Hopkins, A. M., and Sorich, M. J. (2021). Machine Learning for Prediction of Survival Outcomes with Immune-Checkpoint Inhibitors in Urothelial Cancer. *Cancers*, 13(9):2001.

Administration, U. F. a. D. (2025). List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools). *FDA*. Publisher: FDA.

Andres, M. S., Ramalingam, S., Rosen, S. D., Baksi, J., Khattar, R., Kirichenko, Y., Young, K., Yousaf, N., Okines, A., Huddart, R., Harrington, K., Furness, A. J. S., Turajlic, S., Pickering, L., Popat, S., Larkin, J., and Lyon, A. R. (2022). The spectrum of cardiovascular complications related to immune-checkpoint inhibitor treatment : Including myocarditis and the new entity of non inflammatory left ventricular dysfunction. *Cardio-Oncology (London, England)*, 8(1):21.

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., Albright, A., Cheng, J. D., Kang, S. P., Shankaran, V., Piha-Paul, S. A., Yearley, J., Seiwert, T. Y., Ribas, A., and McClanahan, T. K. (2017). IFN—related mRNA profile predicts clinical response to PD-1 blockade. *The Journal of Clinical Investigation*, 127(8):2930–2940.

Balar, A. V., Castellano, D., O'Donnell, P. H., Grivas, P., Vuky, J., Powles, T., Plimack, E. R., Hahn, N. M., de Wit, R., Pang, L., Savage, M. J., Perini, R. F., Keefe, S. M., Bajorin, D., and Bellmunt, J. (2017). First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. *The Lancet Oncology*, 18(11):1483–1492.

- Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., H. Busch, D., Peschel, C., Rad, R., Cox, J., Mann, M., and Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications*, 7(1):13404. Publisher: Nature Publishing Group.
- Benguigui, M., Cooper, T. J., Kalkar, P., Schiff-Zuck, S., Halaban, R., Bacchiocchi, A., Kamer, I., Deo, A., Manobla, B., Menachem, R., Haj-Shomaly, J., Vorontsova, A., Raviv, Z., Buxbaum, C., Christopoulos, P., Bar, J., Lotem, M., Sznol, M., Ariel, A., Shen-Orr, S. S., and Shaked, Y. (2024). Interferon-stimulated neutrophils as a predictor of immunotherapy response. *Cancer Cell*, 42(2):253–265.e12.
- Booth, C. M. and Eisenhauer, E. A. (2012). Progression-free survival: meaningful or simply measurable? *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 30(10):1030–1033.
- Bronger, H., Kraeft, S., Schwarz-Boeger, U., Cerny, C., Stöckel, A., Avril, S., Kiechle, M., and Schmitt, M. (2012). Modulation of CXCR3 ligand secretion by prostaglandin E2 and cyclooxygenase inhibitors in human breast cancer. *Breast cancer research: BCR*, 14(1):R30.
- Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., Johansson, I., Phung, B., Harbst, K., Vallon-Christersson, J., van Schoiack, A., Lövgren, K., Warren, S., Jirström, K., Olsson, H., Pietras, K., Ingvar, C., Isaksson, K., Schadendorf, D., Schmidt, H., Bastholt, L., Carneiro, A., Wargo, J. A., Svane, I. M., and Jönsson, G. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*, 577(7791):561–565. Publisher: Nature Publishing Group.
- Camarena, M. E., Theunissen, P., Ruiz, M., Ruiz-Orera, J., Calvo-Serra, B., Castelo, R., Castro, C., Sarobe, P., Fortes, P., Perera-Bel, J., and



- Albà, M. M. (2024). Microproteins encoded by noncanonical ORFs are a major source of tumor-specific antigens in a liver cancer patient meta-cohort. *Science Advances*, 10(28):eadn3628.
- Cassetta, L. and Pollard, J. W. (2018). Targeting macrophages: therapeutic approaches in cancer. *Nature Reviews Drug Discovery*, 17(12):887–904. Publisher: Nature Publishing Group.
- Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., Kim, J., Kwiatkowski, D. J., Fargo, D. C., Mieczkowski, P. A., Getz, G., and Gordenin, D. A. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nature Genetics*, 47(9):1067–1072.
- Chaput, N., Lepage, P., Coutzac, C., Soularue, E., Le Roux, K., Monot, C., Boselli, L., Routier, E., Cassard, L., Collins, M., Vaysse, T., Marthey, L., Eggermont, A., Asvatourian, V., Lanoy, E., Mateus, C., Robert, C., and Carbonnel, F. (2017). Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab. *Annals of Oncology*, 28(6):1368–1379.
- Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G., and Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*, 11(1):1293.
- Chung, H. C., Ros, W., Delord, J.-P., Perets, R., Italiano, A., Shapira-Frommer, R., Manzuk, L., Piha-Paul, S. A., Xu, L., Zeigenfuss, S., Pruitt, S. K., and Leary, A. (2019). Efficacy and Safety of Pembrolizumab in Previously Treated Advanced Cervical Cancer: Results From the Phase II KEYNOTE-158 Study. *Journal of Clinical Oncology*,

37(17):1470–1478. Publisher: Wolters Kluwer.

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., Sher, X., Liu, X. Q., Lu, H., Nebozhyn, M., Zhang, C., Lunceford, J. K., Joe, A., Cheng, J., Webber, A. L., Ibrahim, N., Plimack, E. R., Ott, P. A., Seiwert, T. Y., Ribas, A., McClanahan, T. K., Tomassini, J. E., Loboda, A., and Kaufman, D. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy. *Science*, 362(6411):eaar3593. Publisher: American Association for the Advancement of Science.

Damrauer, J. S., Beckabir, W., Klomp, J., Zhou, M., Plimack, E. R., Galsky, M. D., Grivas, P., Hahn, N. M., O'Donnell, P. H., Iyer, G., Quinn, D. I., Vincent, B. G., Quale, D. Z., Wobker, S. E., Hoadley, K. A., Kim, W. Y., and Milowsky, M. I. (2022). Collaborative study from the Bladder Cancer Advocacy Network for the genomic analysis of metastatic urothelial cancer. *Nature Communications*, 13(1):6658. Publisher: Nature Publishing Group.

Damsky, W., Jilaveanu, L., Turner, N., Perry, C., Zito, C., Tomayko, M., Leventhal, J., Herold, K., Meffre, E., Bosenberg, M., and Kluger, H. M. (2019). B cell depletion or absence does not impede anti-tumor activity of PD-1 inhibitors. *Journal for ImmunoTherapy of Cancer*, 7(1):153.

Dangaj, D., Bruand, M., Grimm, A. J., Ronet, C., Barras, D., Duttagupta, P. A., Lanitis, E., Duraiswamy, J., Tanyi, J. L., Benencia, F., Conejo-Garcia, J., Ramay, H. R., Montone, K. T., Powell, D. J., Gimotty, P. A., Facciabene, A., Jackson, D. G., Weber, J. S., Rodig, S. J., Hodi, S. F., Kandalaft, L. E., Irving, M., Zhang, L., Foukas, P., Rusakiewicz, S., Delorenzi, M., and Coukos, G. (2019). Cooperation between Constitutive and Inducible Chemokines Enables T-cell Engraftment and Immune Attack in Solid Tumors. *Cancer cell*, 35(6):885–900.e10.

Desbois, M., Udyavar, A. R., Ryner, L., Kozlowski, C., Guan, Y., Dürbaum, M., Lu, S., Fortin, J.-P., Koeppen, H., Ziai, J., Chang, C.-W.,

- Keerthivasan, S., Plante, M., Bourgon, R., Bais, C., Hegde, P., Dae-  
men, A., Turley, S., and Wang, Y. (2020). Integrated digital pathology  
and transcriptome analysis identifies molecular mediators of T-cell  
exclusion in ovarian cancer. *Nature Communications*, 11(1):5583.
- Dhamija, S., Yang, C. M., Seiler, J., Myacheva, K., Caudron-Herger,  
M., Wieland, A., Abdelkarim, M., Sharma, Y., Riester, M., Groß, M.,  
Maurer, J., and Diederichs, S. (2020). A pan-cancer analysis reveals  
nonstop extension mutations causing SMAD4 tumour suppressor  
degradation. *Nature Cell Biology*, 22(8):999–1010. Publisher: Na-  
ture Publishing Group.
- Enkemann, S., Wang, R.-H., Trumbore, M., and Berger, S. (2000).  
Functional discontinuities in prothymosin caused by caspase cleav-  
age apoptotic cells. *Journal of Cellular Physiology*, 182(2):256–268.
- Eroglu, Z., Zaretsky, J. M., Hu-Lieskovan, S., Kim, D. W., Algazi, A.,  
Johnson, D. B., Liniker, E., Ben Kong, Munhoz, R., Rapisuwon, S.,  
Gherardini, P. F., Chmielowski, B., Wang, X., Shintaku, I. P., Wei,  
C., Sosman, J. A., Joseph, R. W., Postow, M. A., Carlino, M. S.,  
Hwu, W.-J., Scolyer, R. A., Messina, J., Cochran, A. J., Long, G. V.,  
and Ribas, A. (2018). High response rate to PD-1 blockade in  
desmoplastic melanomas. *Nature*, 553(7688):347–350. Publisher:  
Nature Publishing Group.
- for Drug Evaluation and Research, C. (2020). FDA approves pem-  
brolizumab for adults and children with TMB-H solid tumors. Pub-  
lisher: FDA.
- Forbes, S. A., Beare, D., Bindal, N., Bamford, S., Ward, S., Cole,  
C. G., Jia, M., Kok, C., Boutselakis, H., De, T., Sondka, Z., Ponting,  
L., Stefancsik, R., Harsha, B., Tate, J., Dawson, E., Thompson, S.,  
Jubb, H., and Campbell, P. J. (2016). COSMIC: High-Resolution  
Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer.  
*Current Protocols in Human Genetics*, 91:10.11.1–10.11.37.

- Gajewski, T. F. (2015). The Next Hurdle in Cancer Immunotherapy: Overcoming the Non-T-Cell-Inflamed Tumor Microenvironment. *Seminars in Oncology*, 42(4):663–671.
- Gejman, R. S., Chang, A. Y., Jones, H. F., DiKun, K., Hakimi, A. A., Schietinger, A., and Scheinberg, D. A. (2018). Rejection of immunogenic tumor clones is limited by clonal fraction. *eLife*, 7:e41090. Publisher: eLife Sciences Publications, Ltd.
- Ghosh, A., Riester, M., Pal, J., Lainde, K.-A., Tangermann, C., Wanning, A., Dueren, U. K., Dhamija, S., and Diederichs, S. (2024). Suppressive cancer nonstop extension mutations increase C-terminal hydrophobicity and disrupt evolutionarily conserved amino acid patterns. *Nature Communications*, 15(1):9209. Publisher: Nature Publishing Group.
- Glaser, A. P., Fantini, D., Wang, Y., Yu, Y., Rimar, K. J., Podojil, J. R., Miller, S. D., and Meeks, J. J. (2017). APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget*, 9(4):4537–4548. Publisher: Impact Journals.
- Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Senn, H.-J., Albain, K. S., André, F., Bergh, J., Bonnefoi, H., Bretel-Morales, D., Burstein, H., Cardoso, F., Castiglione-Gertsch, M., Coates, A. S., Colleoni, M., Costa, A., Curigliano, G., Davidson, N. E., Leo, A. D., Ejlertsen, B., Forbes, J. F., Gelber, R. D., Gnant, M., Goldhirsch, A., Goodwin, P., Goss, P. E., Harris, J. R., Hayes, D. F., Hudis, C. A., Ingle, J. N., Jassem, J., Jiang, Z., Karlsson, P., Loibl, S., Morrow, M., Namer, M., Osborne, C. K., Partridge, A. H., Penault-Llorca, F., Perou, C. M., Piccart-Gebhart, M. J., Pritchard, K. I., Rutgers, E. J. T., Sedlmayer, F., Semiglazov, V., Shao, Z.-M., Smith, I., Thürlimann, B., Toi, M., Tutt, A., Untch, M., Viale, G., Watanabe, T., Wilcken, N., Winer, E. P., and Wood, W. C. (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus

- on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*, 24(9):2206–2223. Publisher: Elsevier.
- Gorbachev, A. V., Kobayashi, H., Kudo, D., Tannenbaum, C. S., Finke, J. H., Shu, S., Farber, J. M., and Fairchild, R. L. (2007). CXC chemokine ligand 9/monokine induced by IFN-gamma production by tumor cells is critical for T cell-mediated suppression of cutaneous tumors. *Journal of Immunology (Baltimore, Md.: 1950)*, 178(4):2278–2286.
- Hammerl, D., Martens, J. W. M., Timmermans, M., Smid, M., Trapman-Jansen, A. M., Foekens, R., Isaeva, O. I., Voorwerk, L., Balcioglu, H. E., Wijers, R., Nederlof, I., Salgado, R., Horlings, H., Kok, M., and Debets, R. (2021). Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nature Communications*, 12(1):5668. Publisher: Nature Publishing Group.
- Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674.
- Harris, D., Haynes, L., Sayles, P., Duso, D., Eaton, S., Lepak, N., Johnson, L., Swain, S., and Lund, F. (2000). Reciprocal regulation of polarized cytokine production by effector B and T cells. *Nature Immunology*, 1(6):475–482.
- Hegde, P. S., Karanikas, V., and Evers, S. (2016). The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. *Clinical Cancer Research*, 22(8):1865–1874.
- Hellmann, M. D., Nathanson, T., Rizvi, H., Creelan, B. C., Sanchez-Vega, F., Ahuja, A., Ni, A., Novik, J. B., Mangarin, L. M. B., Abu-Akeel, M., Liu, C., Sauter, J. L., Rekhtman, N., Chang, E., Callahan, M. K., Chaft, J. E., Voss, M. H., Tenet, M., Li, X.-M., Covello, K., Renninger, A., Vitazka, P., Geese, W. J., Borghaei, H., Rudin, C. M., Antonia, S. J., Swanton, C., Hammerbacher, J., Merghoub, T., McGranahan,

- N., Snyder, A., and Wolchok, J. D. (2018). Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell*, 33(5):843–852.e4. Publisher: Elsevier.
- Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., Yizhak, K., Sade-Feldman, M., Blando, J., Han, G., Gopalakrishnan, V., Xi, Y., Zhao, H., Amaria, R. N., Tawbi, H. A., Cogdill, A. P., Liu, W., LeBleu, V. S., Kugeratski, F. G., Patel, S., Davies, M. A., Hwu, P., Lee, J. E., Gershenwald, J. E., Lucci, A., Arora, R., Woodman, S., Keung, E. Z., Gaudreau, P.-O., Reuben, A., Spencer, C. N., Burton, E. M., Haydu, L. E., Lazar, A. J., Zapassodi, R., Hudgens, C. W., Ledesma, D. A., Ong, S., Bailey, M., Warren, S., Rao, D., Krijgsman, O., Roze-man, E. A., Peeper, D., Blank, C. U., Schumacher, T. N., Butterfield, L. H., Zelazowska, M. A., McBride, K. M., Kalluri, R., Allison, J., Petit-prez, F., Fridman, W. H., Sautès-Fridman, C., Hacohen, N., Rezvani, K., Sharma, P., Tetzlaff, M. T., Wang, L., and Wargo, J. A. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*, 577(7791):549–555.
- Hernando-Calvo, A., Vila-Casadesús, M., Bareche, Y., Gonzalez-Medina, A., Abbas-Aghababazadeh, F., Lo Giacco, D., Martin, A., Saavedra, O., Brana, I., Vieito, M., Fasani, R., Stagg, J., Mancuso, F., Haibe-Kains, B., Han, M., Berche, R., Pugh, T. J., Mirallas, O., Jimenez, J., Gonzalez, N. S., Valverde, C., Muñoz-Couselo, E., Suarez, C., Diez, M., Élez, E., Capdevila, J., Oaknin, A., Saura, C., Macarulla, T., Galceran, J. C., Felip, E., Dienstmann, R., Bedard, P. L., Nuciforo, P., Seoane, J., Tabernero, J., Garralda, E., and Vivancos, A. (2023). A pan-cancer clinical platform to predict immunotherapy outcomes and prioritize immuno-oncology combinations in early-phase trials. *Med*, 4(10):710–727.e5.
- House, I. G., Savas, P., Lai, J., Chen, A. X. Y., Oliver, A. J., Teo, Z. L., Todd, K. L., Henderson, M. A., Giuffrida, L., Petley, E. V., Sek, K., Mardiana, S., Gide, T. N., Quek, C., Scolyer, R. A., Long, G. V., Wilmott, J. S., Loi, S., Darcy, P. K., and Beavis, P. A. (2020). Macrophage-

Derived CXCL9 and CXCL10 Are Required for Antitumor Immune Responses Following Immune Checkpoint Blockade. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 26(2):487–504.

Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hui-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., Seja, E., Lomeli, S., Kong, X., Kelley, M. C., Sosman, J. A., Johnson, D. B., Ribas, A., and Lo, R. S. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell*, 165(1):35–44.

Hussain, K., Kanji, A., Zaheri, S., Malek, D., Terlizzo, M., Weir, J., Turajlic, S., and Fearfield, L. (2023). Checkpoint inhibitor therapy and psoriasis: a case series. *Clinical and Experimental Dermatology*, 48(3):254–256.

Höglund, M., Bernardo, C., Sjö Dahl, G., Eriksson, P., Axelson, H., and Liedberg, F. (2023). The Lund taxonomy for bladder cancer classification - from gene expression clustering to cancer cell molecular phenotypes, and back again. *The Journal of Pathology*, 259(4):369–375.

Jiang, X., Kim, H.-E., Shu, H., Zhao, Y., Zhang, H., Kofron, J., Donnelly, J., Burns, D., Ng, S.-c., Rosenberg, S., and Wang, X. (2003). Distinctive Roles of PHAP Proteins and Prothymosin- in a Death Regulatory Pathway. *Science*, 299(5604):223–226. Publisher: American Association for the Advancement of Science.

Jorgovanovic, D., Song, M., Wang, L., and Zhang, Y. (2020). Roles of IFN- in tumor progression and regression: a review. *Biomarker Research*, 8(1):49.

Jung, J., Heo, Y. J., and Park, S. (2023). High tumor mutational burden predicts favorable response to anti-PD-(L)1 therapy in patients with solid tumor: a real-world pan-tumor analysis. *Journal for Immunotherapy of Cancer*, 11(4):e006454.

- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *Journal of Immunology (Baltimore, Md.: 1950)*, 199(9):3360–3368.
- Kamoun, A., de Reyniès, A., Allory, Y., Sjödaahl, G., Robertson, A. G., Seiler, R., Hoadley, K. A., Groeneveld, C. S., Al-Ahmadie, H., Choi, W., Castro, M. A., Fontugne, J., Eriksson, P., Mo, Q., Kardos, J., Zlotta, A., Hartmann, A., Dinney, C. P., Bellmunt, J., Powles, T., Malats, N., Chan, K. S., Kim, W. Y., McConkey, D. J., Black, P. C., Dyrskjöt, L., Höglund, M., Lerner, S. P., Real, F. X., and Radvanyi, F. (2020). A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *European urology*, 77(4):420–433.
- Keir, M. E., Butte, M. J., Freeman, G. J., and Sharpe, A. H. (2008). PD-1 and Its Ligands in Tolerance and Immunity. *Annual review of immunology*, 26:677–704.
- Kim, J., Kwiatkowski, D., McConkey, D. J., Meeks, J. J., Freeman, S. S., Bellmunt, J., Getz, G., and Lerner, S. P. (2019). The Cancer Genome Atlas Expression Subtypes Stratify Response to Checkpoint Inhibition in Advanced Urothelial Cancer and Identify a Subset of Patients with High Survival Probability. *European Urology*, 75(6):961–964.
- Kim, S. S., Sumner, W. A., Miyauchi, S., Cohen, E. E. W., Califano, J. A., and Sharabi, A. B. (2021). Role of B Cells in Responses to Checkpoint Blockade Immunotherapy and Overall Survival of Cancer Patients. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 27(22):6075–6082.
- Klauer, A. A. and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley interdisciplinary reviews. RNA*, 3(5):649–660.
- Kraemer, A. I., Chong, C., Huber, F., Pak, H., Stevenson, B. J., Müller, M., Michaux, J., Altimiras, E. R., Rusakiewicz, S., Simó-Riudalbas, L.,



- Planet, E., Wiznerowicz, M., Dagher, J., Trono, D., Coukos, G., Tissot, S., and Bassani-Sternberg, M. (2023). The immunopeptidome landscape associated with T cell infiltration, inflammation and immune editing in lung cancer. *Nature Cancer*, 4(5):608–628. Publisher: Nature Publishing Group.
- Lindeboom, R. G. H., Vermeulen, M., Lehner, B., and Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nature Genetics*, 51(11):1645–1651.
- Linnebacher, M., Gebert, J., Rudy, W., Woerner, S., Yuan, Y. P., Bork, P., and von Knebel Doeberitz, M. (2001). Frameshift peptide-derived T-cell epitopes: A source of novel tumor-specific antigens. *International Journal of Cancer*, 93(1):6–11. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.1298>.
- Litchfield, K., Reading, J. L., Lim, E. L., Xu, H., Liu, P., Al-Bakir, M., Wong, Y. N. S., Rowan, A., Funt, S. A., Merghoub, T., Perkins, D., Lauss, M., Svane, I. M., Jönsson, G., Herrero, J., Larkin, J., Quezada, S. A., Hellmann, M. D., Turajlic, S., and Swanton, C. (2020). Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nature Communications*, 11(1):3800. Publisher: Nature Publishing Group.
- Litchfield, K., Reading, J. L., Puttick, C., Thakkar, K., Abbosh, C., Bentham, R., Watkins, T. B., Rosenthal, R., Biswas, D., Rowan, A., Lim, E., Al Bakir, M., Turati, V., Guerra-Assunção, J. A., Conde, L., Furness, A. J., Saini, S. K., Hadrup, S. R., Herrero, J., Lee, S.-H., Van Loo, P., Enver, T., Larkin, J., Hellmann, M. D., Turajlic, S., Quezada, S. A., McGranahan, N., and Swanton, C. (2021). Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell*, 184(3):596–614.e14.
- Long, J., Wang, D., Wang, A., Chen, P., Lin, Y., Bian, J., Yang, X., Zheng, M., Zhang, H., Zheng, Y., Sang, X., and Zhao, H. (2022). A

mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. *Genome Medicine*, 14(1):20.

Maby, P., Galon, J., and Latouche, J.-B. (2016). Frameshift mutations, neoantigens and tumor-specific CD8<sup>+</sup> T cells in microsatellite unstable colorectal cancers. *OncoImmunology*, 5(5):e1115943. Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/2162402X.2015.1115943>.

Malicet, C., Giroux, V., Vasseur, S., Dagorn, J. C., Neira, J. L., and Iovanna, J. L. (2006). Regulation of apoptosis by the p8/prothymosin complex. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8):2671–2676.

Mansfield, A. S., Aubry, M. C., Moser, J. C., Harrington, S. M., Dronca, R. S., Park, S. S., and Dong, H. (2016). Temporal and spatial discordance of programmed cell death-ligand 1 expression and lymphocyte tumor infiltration between paired primary lesions and brain metastases in lung cancer. *Annals of Oncology*, 27(10):1953–1958.

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel III, E. E., Koepfen, H., Astarita, J. L., Cubas, R., Jhunjhunwala, S., Banchereau, R., Yang, Y., Guan, Y., Chalouni, C., Ziai, J., Şenbabaoğlu, Y., Santoro, S., Sheinson, D., Hung, J., Giltane, J. M., Pierce, A. A., Mesh, K., Lianoglou, S., Riegler, J., Carano, R. A. D., Eriksson, P., Höglund, M., Somarriba, L., Halligan, D. L., van der Heijden, M. S., Loriot, Y., Rosenberg, J. E., Fong, L., Mellman, I., Chen, D. S., Green, M., Derleth, C., Fine, G. D., Hegde, P. S., Bourgon, R., and Powles, T. (2018). TGF attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*, 554(7693):544–548. Publisher: Nature Publishing Group.

McGrail, D. J., Pilié, P. G., Rashid, N. U., Voorwerk, L., Slagter, M., Kok, M., Jonasch, E., Khasraw, M., Heimberger, A. B., Lim, B., Ueno, N. T., Litton, J. K., Ferrarotto, R., Chang, J. T., Moulder,

- S. L., and Lin, S.-Y. (2021). High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 32(5):661–672.
- McGranahan, N., Furness, A. J. S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B. K., Shafi, S., Murugaesu, N., Mitter, R., Akarca, A. U., Linares, J., Marafioti, T., Henry, J. Y., Van Allen, E. M., Miao, D., Schilling, B., Schadendorf, D., Garraway, L. A., Makarov, V., Rizvi, N. A., Snyder, A., Hellmann, M. D., Merghoub, T., Wolchok, J. D., Shukla, S. A., Wu, C. J., Peggs, K. S., Chan, T. A., Hadrup, S. R., Quezada, S. A., and Swanton, C. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (New York, N.Y.)*, 351(6280):1463–1469.
- McLaughlin, J., Han, G., Schalper, K. A., Carvajal-Hausdorf, D., Pelekanou, V., Rehman, J., Velcheti, V., Herbst, R., LoRusso, P., and Rimm, D. L. (2016). Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer. *JAMA oncology*, 2(1):46–54.
- Miao, D., Margolis, C. A., Vokes, N. I., Liu, D., Taylor-Weiner, A., Wankowicz, S. M., Adeegbe, D., Keliher, D., Schilling, B., Tracy, A., Manos, M., Chau, N. G., Hanna, G. J., Polak, P., Rodig, S. J., Signoretti, S., Sholl, L. M., Engelman, J. A., Getz, G., Jänne, P. A., Haddad, R. I., Choueiri, T. K., Barbie, D. A., Haq, R., Awad, M. M., Schadendorf, D., Hodi, F. S., Bellmunt, J., Wong, K.-K., Hammerman, P., and Van Allen, E. M. (2018). Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nature genetics*, 50(9):1271–1281.
- Michaeli, J. C., Boch, T., Albers, S., Michaeli, T., and Michaeli, D. T. (2022). Socio-economic burden of disease: Survivorship costs for bladder cancer. *Journal of Cancer Policy*, 32:100326.

- Müller, M., Gfeller, D., Coukos, G., and Bassani-Sternberg, M. (2017). 'Hotspots' of Antigen Presentation Revealed by Human Leukocyte Antigen Ligandomics for Neoantigen Prioritization. *Frontiers in Immunology*, 8. Publisher: Frontiers.
- Nassar, A. H., Mouw, K. W., Jegede, O., Shinagare, A. B., Kim, J., Liu, C.-J., Pomerantz, M., Harshman, L. C., Van Allen, E. M., Wei, X. X., McGregor, B., Choudhury, A. D., Preston, M. A., Dong, F., Signoretti, S., Lindeman, N. I., Bellmunt, J., Choueiri, T. K., Sonpavde, G., and Kwiatkowski, D. J. (2020). A model combining clinical and genomic factors to predict response to PD-1/PD-L1 blockade in advanced urothelial carcinoma. *British Journal of Cancer*, 122(4):555–563. Publisher: Nature Publishing Group.
- Nehrt, N. L., Peterson, T. A., Park, D., and Kann, M. G. (2012). Domain landscapes of somatic mutations in cancer. *BMC Genomics*, 13(Suppl 4):S9.
- Newman, J. H., Chesson, C. B., Herzog, N. L., Bommareddy, P. K., Aspromonte, S. M., Pepe, R., Estupinian, R., Aboelatta, M. M., Budhadev, S., Tarabichi, S., Lee, M., Li, S., Medina, D. J., Giurini, E. F., Gupta, K. H., Guevara-Aleman, G., Rossi, M., Nowicki, C., Abed, A., Goldufsky, J. W., Broucek, J. R., Redondo, R. E., Rotter, D., Jhawar, S. R., Wang, S.-J., Kohlhapp, F. J., Kaufman, H. L., Thomas, P. G., Gupta, V., Kuzel, T. M., Reiser, J., Paras, J., Kane, M. P., Singer, E. A., Malhotra, J., Denzin, L. K., Sant'Angelo, D. B., Rabson, A. B., Lee, L. Y., Lasfar, A., Langenfeld, J., Schenkel, J. M., Fidler, M. J., Ruiz, E. S., Marzo, A. L., Rudra, J. S., Silk, A. W., and Zloza, A. (2020). Intratumoral injection of the seasonal flu shot converts immunologically cold tumors to hot and serves as an immunotherapy for cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 117(2):1119–1128.
- O'Donnell, P. H., Balar, A. V., Vuky, J., Castellano, D. E., Bellmunt, J., Powles, T., Bajorin, D. F., Grivas, P., Hahn, N. M., Plimack, E. R., Savage, M. J., Fang, X., Godwin, J. L., Frenkl, T. L., and

- De Wit, R. (2019). KEYNOTE-052: Phase 2 study evaluating first-line pembrolizumab (pembro) in cisplatin-ineligible advanced urothelial cancer (UC)— Updated response and survival results. *Journal of Clinical Oncology*, 37(15\_suppl):4546–4546. Publisher: Wolters Kluwer.
- O'Donnell, T. J., Rubinsteyn, A., and Laserson, U. (2020). MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Systems*, 11(1):42–48.e7.
- of Cancer Research, F. (2025). TMB Harmonization Project.
- Pal, J., Riester, M., Ganner, A., Ghosh, A., Dhamija, S., Mookherjee, D., Voss, C., Frew, I. J., Kotsis, F., Neumann-Haefelin, E., Spang, A., and Diederichs, S. (2025). Nonstop mutations cause loss of renal tumor suppressor proteins VHL and BAP1 and affect multiple stages of protein translation. *Science Advances*, 11(7):eadr6375. Publisher: American Association for the Advancement of Science.
- Petitprez, F., de Reyniès, A., Keung, E. Z., Chen, T. W.-W., Sun, C.-M., Calderaro, J., Jeng, Y.-M., Hsiao, L.-P., Lacroix, L., Bougoüin, A., Moreira, M., Lacroix, G., Natario, I., Adam, J., Lucchesi, C., Laizet, Y., Toulmonde, M., Burgess, M. A., Bolejack, V., Reinke, D., Wani, K. M., Wang, W.-L., Lazar, A. J., Roland, C. L., Wargo, J. A., Italiano, A., Sautès-Fridman, C., Tawbi, H. A., and Fridman, W. H. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature*, 577(7791):556–560. Publisher: Nature Publishing Group.
- Pond, G. R., Niegisch, G., Rosenberg, J. E., Dreicer, R., Powles, T., Necchi, A., Wei, X. X., Grivas, P., Balar, A. V., Galsky, M. D., Srinivas, S., Choueiri, T. K., Bellmunt, J., Bajorin, D. F., and Sonpavde, G. (2018). New 6-factor prognostic model for patients (pts) with advanced urothelial carcinoma (UC) receiving post-platinum atezolizumab. *Journal of Clinical Oncology*, 36(6\_suppl):413–413. Pub-

lisher: Wolters Kluwer.

- Powles, T., Park, S. H., Voog, E., Caserta, C., Valderrama, B. P., Gurney, H., Kalofonos, H., Radulović, S., Demey, W., Ullén, A., Lorient, Y., Sridhar, S. S., Tsuchiya, N., Kopyltsov, E., Sternberg, C. N., Bellmunt, J., Aragon-Ching, J. B., Petrylak, D. P., Laliberte, R., Wang, J., Huang, B., Davis, C., Fowst, C., Costa, N., Blake-Haskins, J. A., di Pietro, A., and Grivas, P. (2020). Avelumab Maintenance Therapy for Advanced or Metastatic Urothelial Carcinoma. *The New England Journal of Medicine*, 383(13):1218–1230.
- Rech, A. J., Balli, D., Mantero, A., Ishwaran, H., Nathanson, K. L., Stanger, B. Z., and Vonderheide, R. H. (2018). Tumor Immunity and Survival as a Function of Alternative Neopeptides in Human Cancer. *Cancer immunology research*, 6(3):276–287.
- Richman, L. P., Vonderheide, R. H., and Rech, A. J. (2019). Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. *Cell Systems*, 9(4):375–382.e4.
- Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., Lee, W., Yuan, J., Wong, P., Ho, T. S., Miller, M. L., Rekhtman, N., Moreira, A. L., Ibrahim, F., Bruggeman, C., Gasmi, B., Zappasodi, R., Maeda, Y., Sander, C., Garon, E. B., Merghoub, T., Wolchok, J. D., Schumacher, T. N., and Chan, T. A. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (New York, N.Y.)*, 348(6230):124–128.
- Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., Hinoue, T., Laird, P. W., Hoadley, K. A., Akbani, R., Castro, M. A. A., Gibb, E. A., Kanchi, R. S., Gordenin, D. A., Shukla, S. A., Sanchez-Vega, F., Hansel, D. E., Czerniak, B. A., Reuter, V. E., Su, X., de Sa Carvalho, B., Chagas, V. S., Mungall, K. L., Sadeghi, S., Pedamallu, C. S., Lu, Y., Klimczak, L. J., Zhang, J., Choo, C.,

- Ojesina, A. I., Bullman, S., Leraas, K. M., Lichtenberg, T. M., Wu, C. J., Schultz, N., Getz, G., Meyerson, M., Mills, G. B., McConkey, D. J., TCGA Research Network, Weinstein, J. N., Kwiatkowski, D. J., and Lerner, S. P. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell*, 171(3):540–556.e25.
- Robertson, A. G., Meghani, K., Cooley, L. F., McLaughlin, K. A., Fall, L. A., Yu, Y., Castro, M. A. A., Groeneveld, C. S., de Reyniès, A., Nazarov, V. I., Tsvetkov, V. O., Choy, B., Raggi, D., Marandino, L., Montorsi, F., Powles, T., Necchi, A., and Meeks, J. J. (2023). Expression-based subtypes define pathologic response to neoadjuvant immune-checkpoint inhibitors in muscle-invasive bladder cancer. *Nature Communications*, 14(1):2126. Publisher: Nature Publishing Group.
- Rogel, A., Vignard, V., Bobinet, M., Labarriere, N., and Lang, F. (2011). A long peptide from MELOE-1 contains multiple HLA class II T cell epitopes in addition to the HLA-A\*0201 epitope: an attractive candidate for melanoma vaccination. *Cancer immunology, immunotherapy: CII*, 60(3):327–337.
- Rose, T. L., Weir, W. H., Mayhew, G. M., Shibata, Y., Eulitt, P., Uronis, J. M., Zhou, M., Nielsen, M., Smith, A. B., Woods, M., Hayward, M. C., Salazar, A. H., Milowsky, M. I., Wobker, S. E., McGinty, K., Millburn, M. V., Eisner, J. R., and Kim, W. Y. (2021). Fibroblast growth factor receptor 3 alterations and response to immune checkpoint inhibition in metastatic urothelial cancer: a real world experience. *British Journal of Cancer*, 125(9):1251–1260.
- Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., Dawson, N., O'Donnell, P. H., Balmanoukian, A., Loriot, Y., Srinivas, S., Retz, M. M., Grivas, P., Joseph, R. W., Galsky, M. D., Fleming, M. T., Petrylak, D. P., Perez-Gracia, J. L., Burris, H. A., Castellano, D., Canil, C., Bellmunt, J., Bajorin, D., Nickles, D., Bourgon, R., Frampton, G. M., Cui, N., Mariathasan, S., Abidoye, O., Fine, G. D., and Dreicer, R. (2016). Atezolizumab

in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet*, 387(10031):1909–1920.

Rosenthal, R., Cadieux, E. L., Salgado, R., Bakir, M. A., Moore, D. A., Hiley, C. T., Lund, T., Tanić, M., Reading, J. L., Joshi, K., Henry, J. Y., Ghorani, E., Wilson, G. A., Birkbak, N. J., Jamal-Hanjani, M., Veeriah, S., Szallasi, Z., Loi, S., Hellmann, M. D., Feber, A., Chain, B., Herrero, J., Quezada, S. A., Demeulemeester, J., Van Loo, P., Beck, S., McGranahan, N., and Swanton, C. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature*, 567(7749):479–485. Publisher: Nature Publishing Group.

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., Barron, D. A., Zehir, A., Jordan, E. J., Omuro, A., Kaley, T. J., Kendall, S. M., Motzer, R. J., Hakimi, A. A., Voss, M. H., Russo, P., Rosenberg, J., Iyer, G., Bochner, B. H., Bajorin, D. F., Al-Ahmadie, H. A., Chaft, J. E., Rudin, C. M., Riely, G. J., Baxi, S., Ho, A. L., Wong, R. J., Pfister, D. G., Wolchok, J. D., Barker, C. A., Gutin, P. H., Brennan, C. W., Tabar, V., Mellingerhoff, I. K., DeAngelis, L. M., Ariyan, C. E., Lee, N., Tap, W. D., Gounder, M. M., D'Angelo, S. P., Saltz, L., Stadler, Z. K., Scher, H. I., Baselga, J., Razavi, P., Klebanoff, C. A., Yaeger, R., Segal, N. H., Ku, G. Y., DeMatteo, R. P., Ladanyi, M., Rizvi, N. A., Berger, M. F., Riaz, N., Solit, D. B., Chan, T. A., and Morris, L. G. T. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics*, 51(2):202–206. Publisher: Nature Publishing Group.

Schaafsma, E., Jiang, C., and Cheng, C. (2021). B cell infiltration is highly associated with prognosis and an immune-infiltrated tumor microenvironment in neuroblastoma. *Journal of cancer metastasis and treatment*, 7(34):10.20517/2394-4722.2021.72.

Schmid, P., Adams, S., Rugo, H. S., Schneeweiss, A., Barrios, C. H., Iwata, H., Diéras, V., Hegg, R., Im, S.-A., Wright,



- G. S., Henschel, V., Molinero, L., Chui, S. Y., Funke, R., Husain, A., Winer, E. P., Loi, S., and Emens, L. A. (2018). Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *New England Journal of Medicine*, 379(22):2108–2121. Publisher: Massachusetts Medical Society \_eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1809615>.
- Shapiro, I. E. and Bassani-Sternberg, M. (2023). The impact of immuno-peptidomics: From basic research to clinical implementation. *Seminars in Immunology*, 66:101727.
- Shibata, N., Ohoka, N., Sugaki, Y., Onodera, C., Inoue, M., Sakuraba, Y., Takakura, D., Hashii, N., Kawasaki, N., Gondo, Y., and Naito, M. (2015). Degradation of Stop Codon Read-through Mutant Proteins via the Ubiquitin-Proteasome System Causes Hereditary Disorders. *The Journal of Biological Chemistry*, 290(47):28428–28437.
- Snyder, A., Nathanson, T., Funt, S. A., Ahuja, A., Buros Novik, J., Hellmann, M. D., Chang, E., Aksoy, B. A., Al-Ahmadie, H., Yuskov, E., Vignali, M., Benzeno, S., Boyd, M., Moran, M., Iyer, G., Robins, H. S., Mardis, E. R., Merghoub, T., Hammerbacher, J., Rosenberg, J. E., and Bajorin, D. F. (2017). Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. *PLoS medicine*, 14(5):e1002309.
- Sullivan, R. J. and Weber, J. S. (2022). Immune-related toxicities of checkpoint inhibitors: mechanisms and mitigation strategies. *Nature Reviews Drug Discovery*, 21(7):495–508. Publisher: Nature Publishing Group.
- Swanton, C., McGranahan, N., Starrett, G. J., and Harris, R. S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discovery*, 5(7):704–712.
- Togashi, Y., Shitara, K., and Nishikawa, H. (2019). Regulatory T cells in cancer immunosuppression — implications for anticancer therapy.

*Nature Reviews Clinical Oncology*, 16(6):356–371.

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J. L., Wong, Y. N. S., Rowan, A., Kanu, N., Al Bakir, M., Chambers, T., Salgado, R., Savas, P., Loi, S., Birkbak, N. J., Sansregret, L., Gore, M., Larkin, J., Quezada, S. A., and Swanton, C. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *The Lancet. Oncology*, 18(8):1009–1021.

Usset, J., Rosendahl Huber, A., Andrianova, M. A., Batlle, E., Carles, J., Cuppen, E., Elez, E., Felip, E., Gómez-Rey, M., Lo Giacco, D., Martinez-Jimenez, F., Muñoz-Couselo, E., Siu, L. L., Tabernero, J., Vivancos, A., Muiños, F., Gonzalez-Perez, A., and Lopez-Bigas, N. (2024). Five latent factors underlie response to immunotherapy. *Nature Genetics*, 56(10):2112–2120. Publisher: Nature Publishing Group.

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Geukes Foppen, M. H., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., Wu, C. J., Schadendorf, D., and Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, 350(6257):207–211. Publisher: American Association for the Advancement of Science.

Viaud, S., Saccheri, F., Mignot, G., Yamazaki, T., Daillère, R., Hannani, D., Enot, D. P., Pfirschke, C., Engblom, C., Pittet, M. J., Schlitzer, A., Ginhoux, F., Apetoh, L., Chachaty, E., Woerther, P.-L., Eberl, G., Bérard, M., Ecobichon, C., Clermont, D., Bizet, C., Gaboriau-Routhiau, V., Cerf-Bensussan, N., Opolon, P., Yessaad, N., Vivier, E., Ryffel, B., Elson, C. O., Doré, J., Kroemer, G., Lepage, P., Boneca, I. G., Ghiringhelli, F., and Zitvogel, L. (2013). The Intestinal Microbiota Modulates the Anticancer Immune Effects of Cyclophosphamide. *Science*, 342(6161):971–976. Publisher: American Association for

the Advancement of Science.

Vétizou, M., Pitt, J. M., Daillère, R., Lepage, P., Waldschmitt, N., Flament, C., Rusakiewicz, S., Routy, B., Roberti, M. P., Duong, C. P. M., Poirier-Colame, V., Roux, A., Becharef, S., Formenti, S., Golden, E., Cording, S., Eberl, G., Schlitzer, A., Ginhoux, F., Mani, S., Yamazaki, T., Jacquelot, N., Enot, D. P., Bérard, M., Nigou, J., Opolon, P., Eggermont, A., Woerther, P.-L., Chachaty, E., Chaput, N., Robert, C., Mateus, C., Kroemer, G., Raoult, D., Boneca, I. G., Carbonnel, F., Chamaillard, M., and Zitvogel, L. (2015). Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science*, 350(6264):1079–1084. Publisher: American Association for the Advancement of Science.

Wang, J. E. and Livingstone, A. M. (2003). Cutting Edge: CD4+ T Cell Help Can Be Essential for Primary CD8+ T Cell Responses In Vivo 1. *The Journal of Immunology*, 171(12):6339–6343.

Wang, L., Saci, A., Szabo, P. M., Chasalow, S. D., Castillo-Martin, M., Domingo-Domenech, J., Siefker-Radtke, A., Sharma, P., Sfakianos, J. P., Gong, Y., Dominguez-Andres, A., Oh, W. K., Mulholland, D., Azrilevich, A., Hu, L., Cordon-Cardo, C., Salmon, H., Bhardwaj, N., Zhu, J., and Galsky, M. D. (2018). EMT- and stroma-related gene expression and resistance to PD-1 blockade in urothelial cancer. *Nature Communications*, 9:3503.

Wang, Y., Kim, T. H., Fouladdel, S., Zhang, Z., Soni, P., Qin, A., Zhao, L., Azizi, E., Lawrence, T. S., Ramnath, N., Cuneo, K. C., and Nagrath, S. (2019). PD-L1 Expression in Circulating Tumor Cells Increases during Radio(chemo)therapy and Indicates Poor Prognosis in Non-small Cell Lung Cancer. *Scientific Reports*, 9(1):566. Publisher: Nature Publishing Group.

Weinstein, J. N., Akbani, R., Broom, B. M., Wang, W., Verhaak, R. G. W., McConkey, D., Lerner, S., Morgan, M., Creighton, C. J., Smith, C., Kwiatkowski, D. J., Cherniack, A. D., Kim, J., Sekhar Pedamallu,

C., Noble, M. S., Al-Ahmadie, H. A., Reuter, V. E., Rosenberg, J. E., Bajorin, D. F., Bochner, B. H., Solit, D. B., Koppie, T., Robinson, B., Gordenin, D. A., Fargo, D., Klimczak, L. J., Roberts, S. A., Au, J., Laird, P. W., Hinoue, T., Schultz, N., Ramirez, R., Hansel, D., Hoadley, K. A., Kim, W. Y., Damrauer, J. S., Baylin, S. B., Mungall, A. J., Gordon Robertson, A., Chu, A., Kwiatkowski, D. J., Sougnez, C., Cibulskis, K., Lichtenstein, L., Sivachenko, A., Stewart, C., Lawrence, M. S., Getz, G., Lander, E., Gabriel, S. B., Creighton, C. J., Donehower, L., Cherniack, A. D., Kim, J., Carter, S. L., Sak-sena, G., Schumacher, S. E., Sougnez, C., Freeman, S. S., Jung, J., Sekhar Pedamallu, C., Bhatt, A. S., Pugh, T., Getz, G., Beroukhim, R., Gabriel, S. B., Meyerson, M., Mungall, A. J., Gordon Robertson, A., Chu, A., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Dhalla, N., Hirst, C., Holt, R. A., Jones, S. J. M., Lee, D., Li, H. I., Marra, M. A., Mayo, M., Moore, R. A., Schein, J. E., Sipahimalani, P., Tam, A., Thiessen, N., Wong, T., Wye, N., Bowlby, R., Chuah, E., Guin, R., Jones, S. J. M., Marra, M. A., Hinoue, T., Shen, H., Bootwalla, M. S., Triche Jr, T., Lai, P. H., Van Den Berg, D. J., Weisenberger, D. J., Laird, P. W., Hansel, D., Hoadley, K. A., Balu, S., Bodenheimer, T., Damrauer Alan P. Hoyle, J. S., Jefferys, S. R., Meng, S., Mose, L. E., Simons, J. V., Soloway, M. G., Wu, J., Kim, W. Y., Parker, J. S., Neil Hayes, D., Roach, J., Buda, E., Jones, C. D., Mieczkowski, P. A., Tan, D., Veluvolu, U., Waring, S., Todd Auman, J., Perou, C. M., Wilkerson, M. D., Santoso, N., Parfenov, M., Ren, X., Pantazi, A., Hadjipanayis, A., Seidman, J., Kucherlapati, R., Lee, S., Yang, L., Park, P. J., Baylin, S. B., Wei Xu, A., Protopopov, A., Zhang, J., Bristow, C., Mahadeshwar, H. S., Seth, S., Song, X., Tang, J., Zeng, D., Chin, L., Guo, C., Weinstein, J. N., Akbani, R., Broom, B. M., McConkey, D., Casasent, T. D., Liu, W., Ju, Z., Motter, T., Peng, B., Ryan, M., Wang, W., Verhaak, R. G. W., Su, X., Yang, J.-Y., Lorenzi, P. L., Yao, H., Zhang, N., Zhang, J., Mills, G. B., Kim, J., Noble, M. S., Cho, J., DiCara, D., Frazer, S., Gehlenborg, N., Heiman, D. I., Lin, P., Liu, Y., Stojanov, P., Voet, D., Zhang, H., Zou, L., Chin, L., Getz, G., Bernard, B., Kreisberg, D., Reynolds, S., Rovira,

H., Shmulevich, I., Ramirez, R., Schultz, N., Gao, J., Jacobsen, A., Arman Aksoy, B., Antipin, Y., Ciriello, G., Dresdner, G., Gross, B., Lee, W., Reva, B., Shen, R., Sinha, R., Onur Sumer, S., Weinhold, N., Ladanyi, M., Sander, C., Benz, C., Carlin, D., Haussler, D., Ng, S., Paull, E. O., Stuart, J., Zhu, J., Liu, Y., Zhang, W., Taylor, B. S., Lichtenberg, T. M., Zmuda, E., Barr, T., Black, A. D., George, M., Hanf, B., Helsel, C., McAllister, C., Ramirez, N. C., Tabler, T. R., Weaver, S., Wise, L., Bowen, J., Gastier-Foster, J. M., Weinstein, J. N., Lerner, S., Jian, W., Tello, S., Ittman, M., Castro, P., McClenden, W. D., Morgan, M., Gibbs, R., Liu, Y., Saller, C., Tarvin, K., DiPiero, J. M., Owens, J., Bollag, R., Li, Q., Weinberger, P., Czerwinski, C., Huelsenbeck-Dill, L., Iacocca, M., Petrelli, N., Rabeno, B., The Cancer Genome Atlas Research Network, Analysis working group: The University of Texas MD Anderson Cancer Center, Baylor College of Medicine, Broad Institute, Memorial Sloan-Kettering Cancer Center, Oregon Health and Science University, D. o. U., Weill Medical College of Cornell University, National Institute of Environmental Health Sciences, Optimum Therapeutics LLC, University of Southern California Epigenome Center, Computational Biology Center, M. S.-K. C. C., UCSD Department of Pathology, Lineberger Comprehensive Cancer Center, U. o. N. C. a. C. H., Department of Genetics, U. o. N. C. a. C. H., The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Canada's Michael Smith Genome Sciences Centre, B. C. A., Genome Sequencing Center: Broad Institute, Genome characterization centres: Dan L. Duncan Cancer Center, Baylor College of Medicine, H. G. S. C., Canada's Michael Smith Genome Sciences Centre, B. C. A., Research Computing Center, U. o. N. C. a. C. H., Carolina Center for Genome Sciences, U. o. N. C. a. C. H., Department of Biology, U. o. N. C. a. C. H., Eshelman School of Pharmacy, U. o. N. C. a. C. H., Department of Genetics, H. M. S., The Center for Biomedical Informatics, H. M. S., Cancer Biology Division, T. S. K. C. C. C. a. J. H. U., Division of Genetics, B. a. W. H., Institute for Applied Cancer Science, The University of Texas MD Anderson Cancer Center, D. o. G. M., The University of

Texas MD Anderson Cancer Center, D. o. P., Genome data analysis centres: The University of Texas M.D. Anderson Cancer Center, Institute for Systems Biology, Buck Institute for Research on Aging, University of California Santa Cruz, Department of Pathology, M. A. C. C., Helen Diller Family Comprehensive Cancer Center, U. o. C., Biospecimen core resource: The Research Institute at Nationwide Children's Hospital, Tissue source sites: The University of Texas MD Anderson Cancer Center, Scott Department of Urology, B. C. o. M., Texas Cancer Research Biobank (TCRB), B. C. o. M., Analytical Biological Services, I., Cleveland Clinic Foundation, Georgia Regents University Cancer Center, and Helen F. Graham Cancer Center at Christiana Care (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, 507(7492):315–322. Publisher: Nature Publishing Group.

Wojtukiewicz, M. Z., Rek, M. M., Karpowicz, K., Górská, M., Polityńska, B., Wojtukiewicz, A. M., Moniuszko, M., Radziwon, P., Tucker, S. C., and Honn, K. V. (2021). Inhibitors of immune checkpoints—PD-1, PD-L1, CTLA-4—new opportunities for cancer patients and a new challenge for internists and general practitioners. *Cancer Metastasis Reviews*, 40(3):949–982.

Wolf, Y., Bartok, O., Patkar, S., Eli, G. B., Cohen, S., Litchfield, K., Levy, R., Jiménez-Sánchez, A., Trabish, S., Lee, J. S., Karathia, H., Barnea, E., Day, C.-P., Cinnamon, E., Stein, I., Solomon, A., Bitton, L., Pérez-Guijarro, E., Dubovik, T., Shen-Orr, S. S., Miller, M. L., Merlino, G., Levin, Y., Pikarsky, E., Eisenbach, L., Admon, A., Swanton, C., Ruppin, E., and Samuels, Y. (2019). UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. *Cell*, 179(1):219–235.e21. Publisher: Elsevier.

Yadav, M., Jhunjhunwala, S., Phung, Q. T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., Cheung, T. K., Fritsche, J., Weinschenk, T., Modrusan, Z., Mellman, I., Lill, J. R., and Delamarre, L. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, 515(7528):572–576. Pub-

lisher: Nature Publishing Group.

Yin, Q., Wu, L., Han, L., Zheng, X., Tong, R., Li, L., Bai, L., and Bian, Y. (2023). Immune-related adverse events of immune checkpoint inhibitors: a review. *Frontiers in Immunology*, 14:1167975.

Yoo, S.-K., Fitzgerald, C. W., Cho, B. A., Fitzgerald, B. G., Han, C., Koh, E. S., Pandey, A., Sfreddo, H., Crowley, F., Korostin, M. R., Debnath, N., Leyfman, Y., Valero, C., Lee, M., Vos, J. L., Lee, A. S., Zhao, K., Lam, S., Olumuyide, E., Kuo, F., Wilson, E. A., Hamon, P., Hennequin, C., Saffern, M., Vuong, L., Hakimi, A. A., Brown, B., Merad, M., Gnjjatic, S., Bhardwaj, N., Galsky, M. D., Schadt, E. E., Samstein, R. M., Marron, T. U., Gönen, M., Morris, L. G. T., and Chowell, D. (2025). Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data. *Nature Medicine*, pages 1–12. Publisher: Nature Publishing Group.

Zhang, Q.-F., Li, J., Jiang, K., Wang, R., Ge, J.-L., Yang, H., Liu, S.-J., Jia, L.-T., Wang, L., and Chen, B.-L. (2020). CDK4/6 inhibition promotes immune infiltration in ovarian cancer and synergizes with PD-1 blockade in a B cell-dependent manner. *Theranostics*, 10(23):10619–10633.

Zhang, X., Kim, S., Hundal, J., Herndon, J. M., Li, S., Petti, A. A., Soysal, S. D., Li, L., McLellan, M. D., Hoog, J., Primeau, T., Myers, N., Vickery, T. L., Sturmoski, M., Hagemann, I. S., Miller, C. A., Ellis, M. J., Mardis, E. R., Hansen, T., Fleming, T. P., Goedegebuure, S. P., and Gillanders, W. E. (2017). Breast Cancer Neoantigens Can Induce CD8+ T-Cell Responses and Antitumor Immunity. *Cancer Immunology Research*, 5(7):516–523.

Zhang, Z., Wang, Z.-X., Chen, Y.-X., Wu, H.-X., Yin, L., Zhao, Q., Luo, H.-Y., Zeng, Z.-L., Qiu, M.-Z., and Xu, R.-H. (2022). Integrated analysis of single-cell and bulk RNA sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response. *Genome Medicine*, 14(1):45.





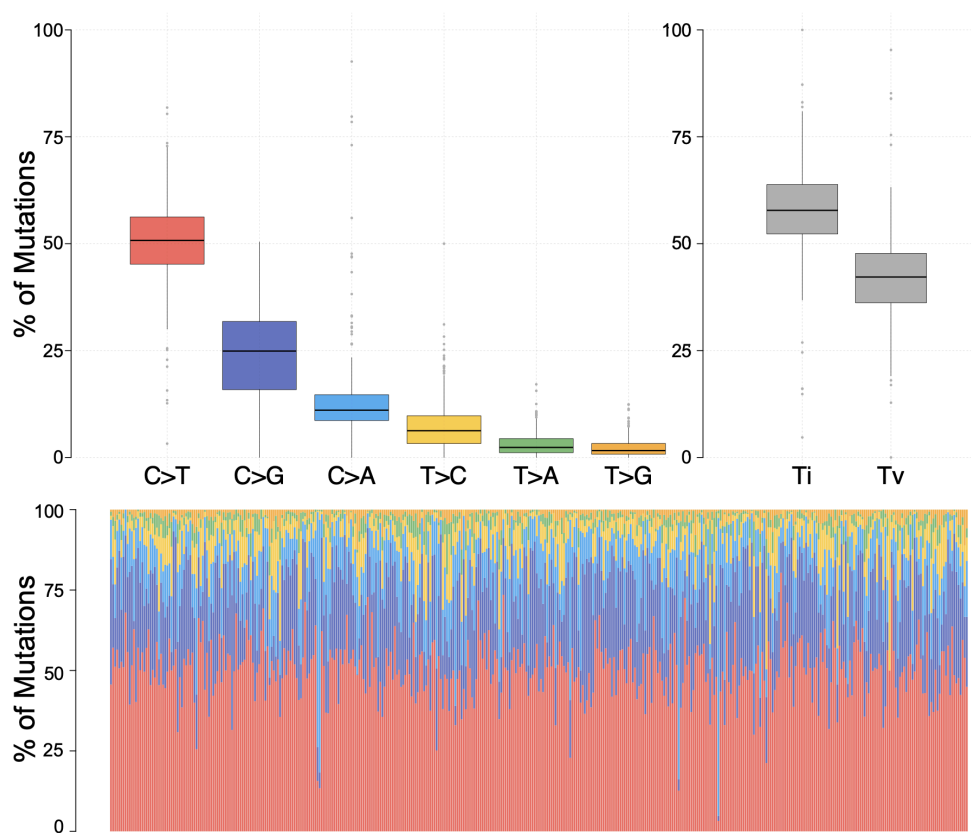
## SUPPLEMENTARY MATERIAL

### 5.1 Supplementary Material for Results 3.1

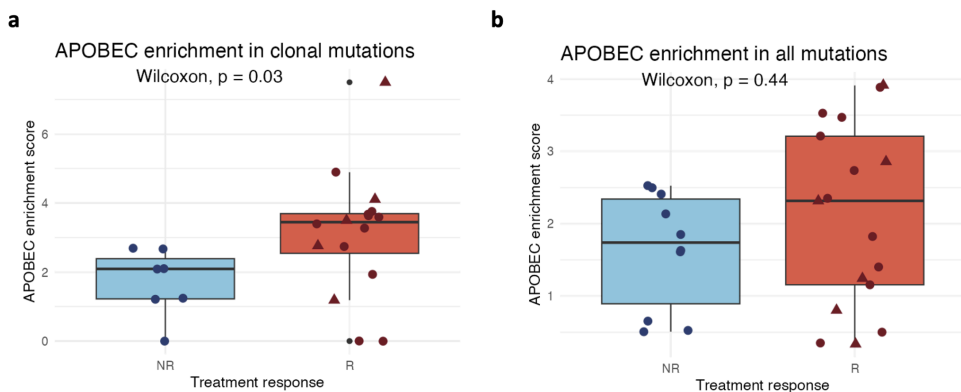
Supplementary figures and tables for the publication titled 'The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer', presented in section 3.1.

**Full citation:**

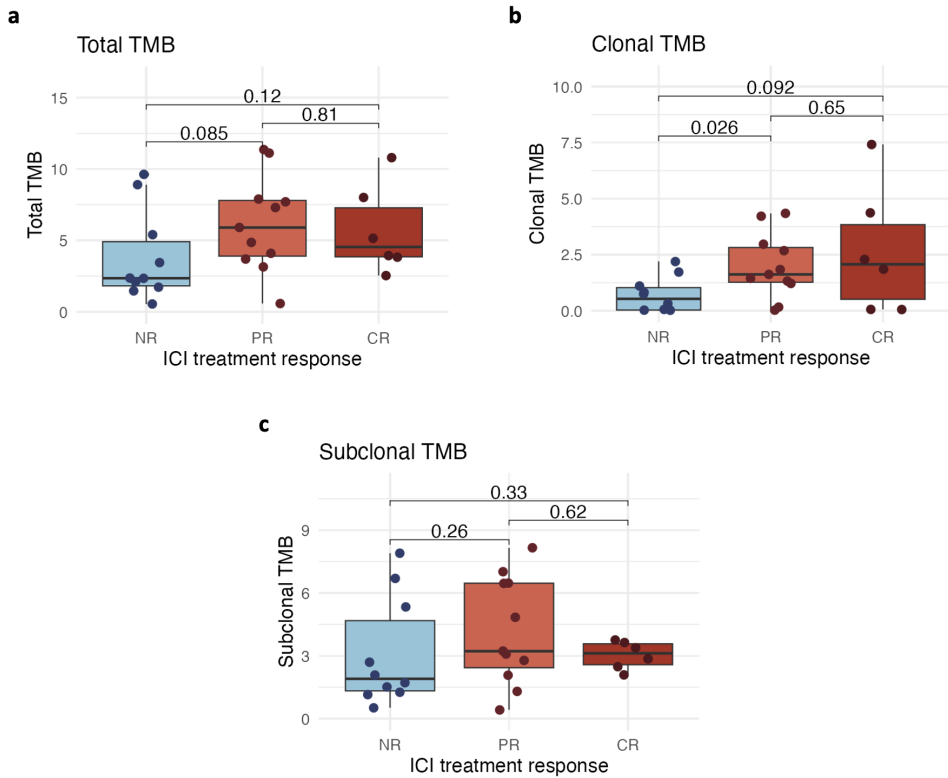
Boll, L.M., Perera-Bel, J., Rodriguez-Vida, A. et al. The impact of mutational clonality in predicting the response to immune checkpoint inhibitors in advanced urothelial cancer. Sci Rep 13, 15287 (2023). <https://doi.org/10.1038/s41598-023-42495-2>.



**Figure S1. Pairwise nucleotide substitution patterns in the urothelial cancer cohort of the Cancer Genome Atlas (TCGA).** The frequency of transitions (Ti) and transversions (Tv) is also shown. The number of samples is 411.

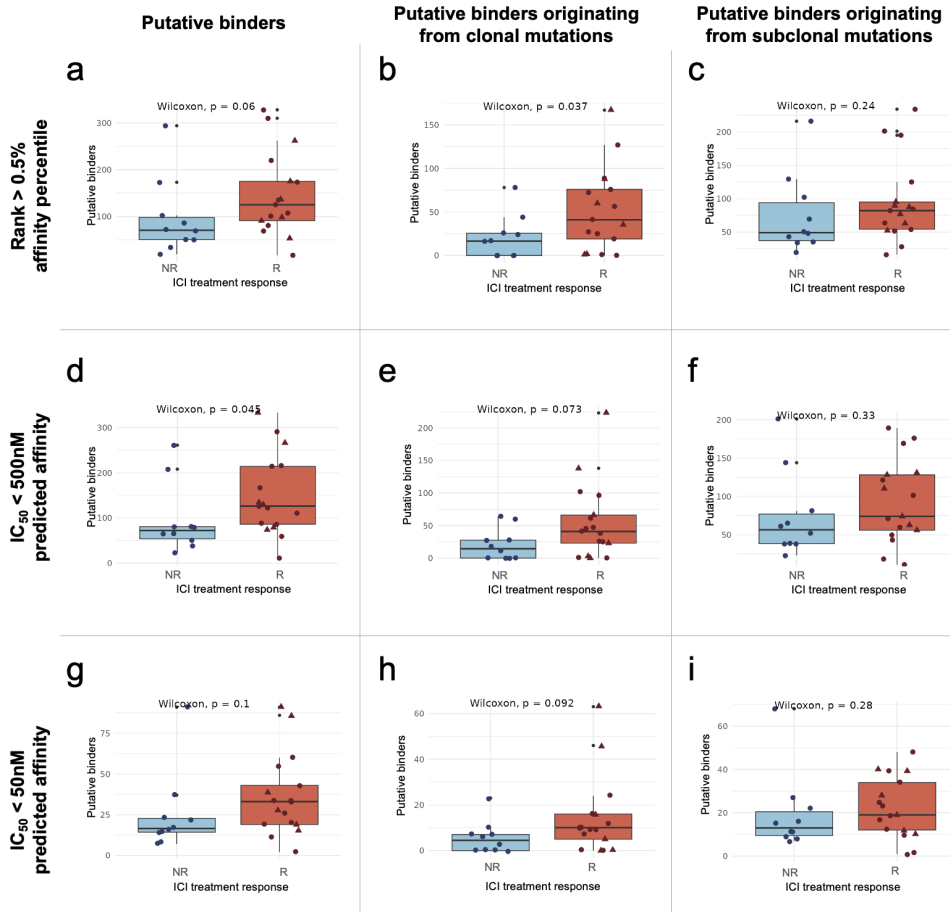


**Figure S2. Relationship between APOBEC mutations and response to treatment. a. APOBEC enrichment in clonal mutations.** There is a significant positive relationship between response to treatment and the APOBEC enrichment in clonal mutations ( $p$ -value = 0.03, Wilcoxon rank sum test). **b. APOBEC enrichment in all mutations.** The differences between responders and non-responders are not statistically significant. NR: no responders, R: responders, triangle shape represents the complete responders among the responder group.



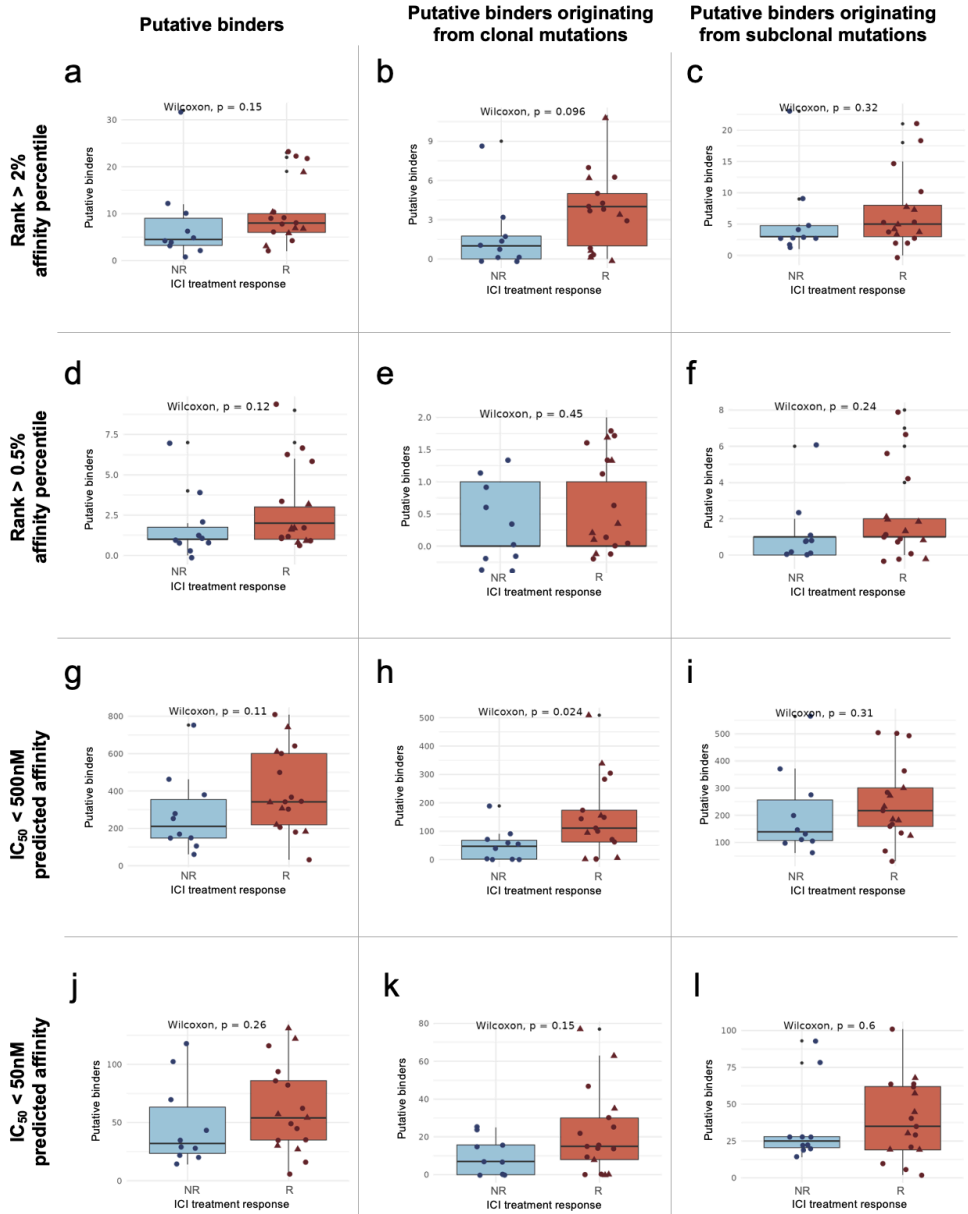
**Figure S3. Relationship between TMB and the three different response groups to ICI therapy. a. Relationship between TMB and response to ICI treatment.** The differences in TMB values between the three response groups were not significant. **b. Relationship between clonalTMB and response to ICI treatment.** The clonal TMB for partial responders is significantly higher compared to non-responders (p-value = 0.026). The differences in clonal TMB between non-responders and complete responders and partial responders compared to complete responders did not reach statistical significance. **c. Relationship between subclonal TMB and response to ICI treatment.** The differences in subclonal TMB values between the three response groups were not significant. NR: no responders, PR: partial responders, CR: complete responders.

# NetMHCpan 4.0

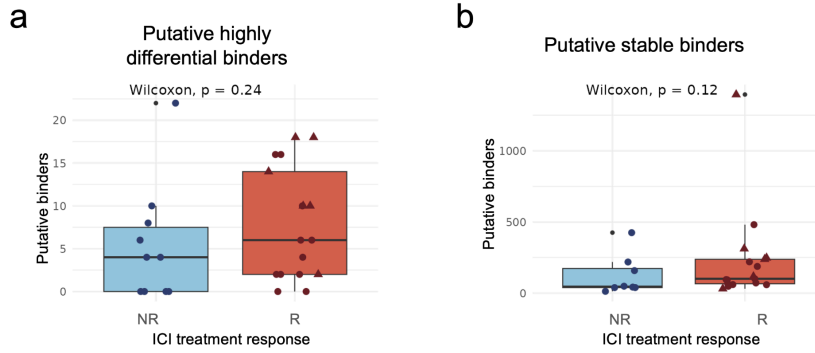


**Figure S4. Relationship between ICI treatment response and the number of putative binders predicted with NetMHCpan 4.0.** The number of predicted binders tends to be higher in responders than in non-responders. **a-c:** Number of putative binders with predicted affinity rank < 0.5%. **d-f:** Number of putative binders with predicted affinity  $IC_{50} < 500nM$ . **g-i:** Number of putative binders with predicted affinity  $IC_{50} < 50nM$ . P-values were calculated using the Wilcoxon rank sum test. NR: no responders, R: responders, triangle shape represents the complete responders among the responder group.

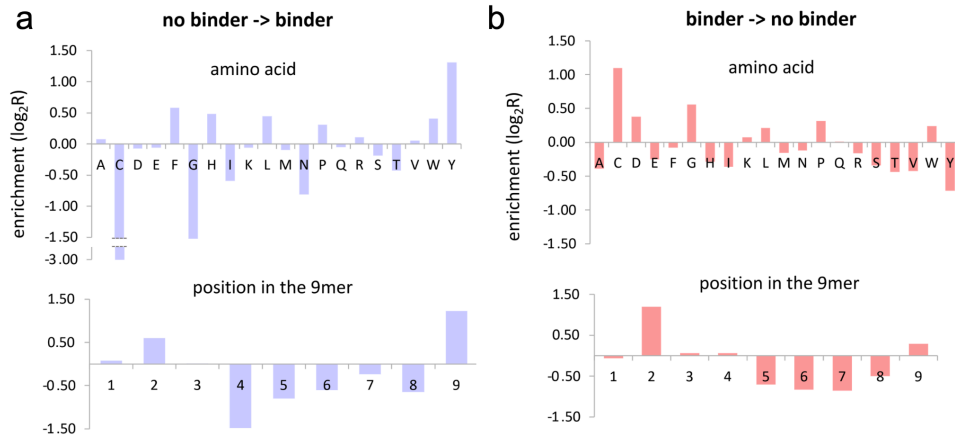
# MHCflurry 2.0



**Figure S5. Relationship between ICI treatment response and the number of putative binders predicted with MHCflurry 2.0.** The number of putative binders tends to be higher in responders than non-responders. **a-c:** Number of putative binders with predicted affinity rank < 2%. **d-f:** Number of putative binders with predicted affinity rank < 0.5%. **g-i:** Number of putative binders with predicted affinity  $IC_{50} < 500nM$ . **j-l:** Number of putative binders with predicted affinity  $IC_{50} < 50nM$ . P-values were calculated using the Wilcoxon rank sum test. NR: no responders, R: responders, triangle shape represents the complete responders among the responder group.

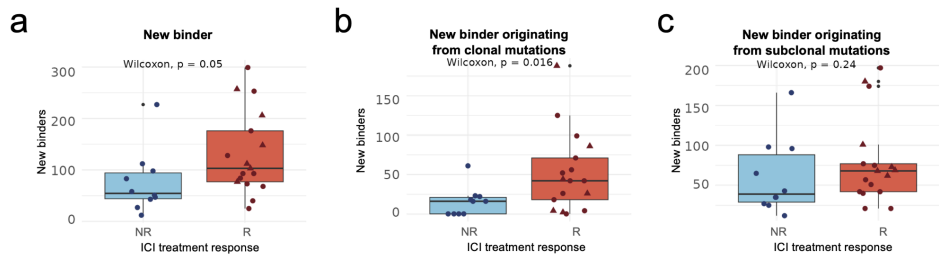


**Figure S6. a. Relationship of differential agretopicity index (DAI) and treatment response.** No significant difference in highly differential peptides (DAI>9) was seen between response groups. **b. Relationship between the number of predicted stable binders and treatment response.** No significant difference in the number of putative stable binders (binding stability<1.4h) was seen between response groups. P-values were calculated using Wilcoxon rank sum test. NR: no responders, R: responders, triangle shape represents the complete responders among the responder group.



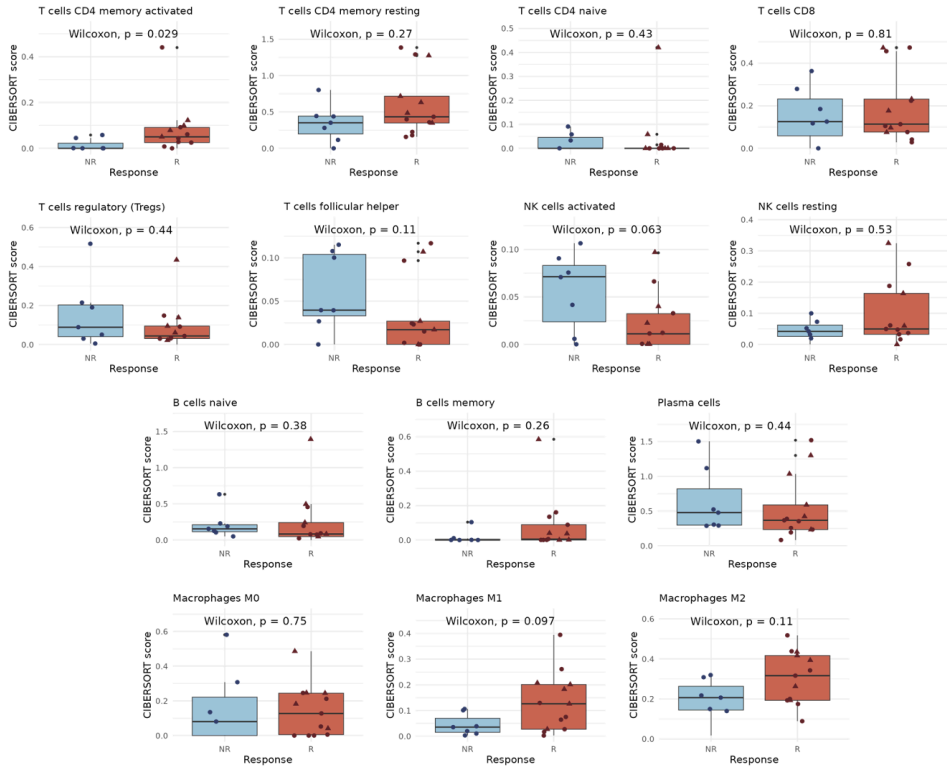
**Figure S7. Non-synonymous mutations affect the binding affinity of the peptide. a. Formation of new MHC I binders.** Enrichment for different amino acid substitutions (above) or positions in the peptide (below) are measured as the log<sub>2</sub> ratio of the substitution frequency in the set of new binders versus the frequency in the set of peptides that do not change binding status. Enriched amino acids using a chi-square test: tyrosine (Y), phenylalanine (F), leucine (L) and histidine (H) at p-value < 10<sup>-5</sup>, tryptophan (W) at p-value = 0.002872. **b. Loss of MHC I binding capacity.** Enrichment for different amino acid substitutions (above) or positions in the peptide (below) are measured as the log<sub>2</sub> ratio of the substitution frequency in the set of peptides associated with loss of MHC I binding *versus* the frequency in the set of peptides that do not change their binding status. Enriched amino acids using a chi-square test: cysteine (C) (p-value < 10<sup>-5</sup>), glycine (G) (p-value = 6.55x10<sup>-5</sup>).



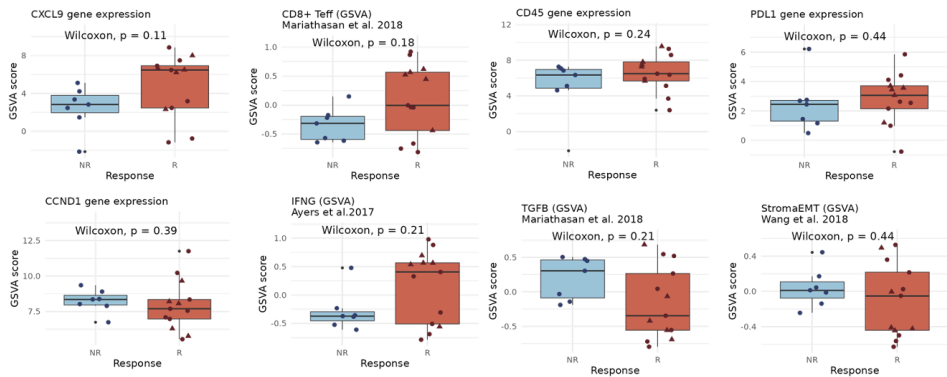


**Figure S8. Relationship of number of predicted new binders and treatment response. a. Binders derived from total mutations.** Responders have a higher number of new putative binders than non-responders (Wilcoxon test, p-value = 0.05). **b. Binders derived from clonal mutations.** Number of putative binders originating from clonal mutations is significantly higher in responders than non-responders (Wilcoxon test, p-value = 0.016). **c. Binders derived from subclonal mutations.** No significant difference can be observed for the number of putative binders originating from subclonal mutations between responders and non-responders. NR: no responders, R: responders, triangle shape represents the complete responders among the responder group.

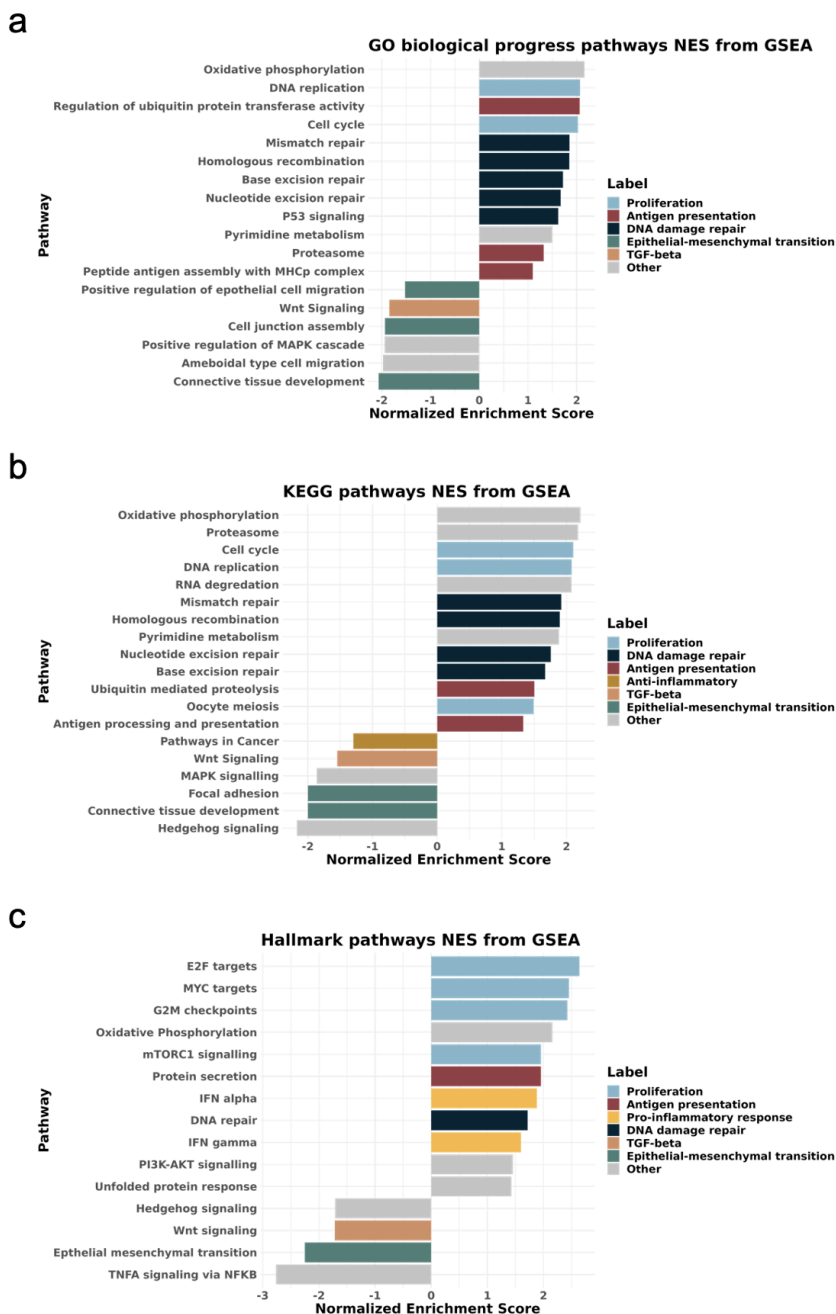
**a**



**b**

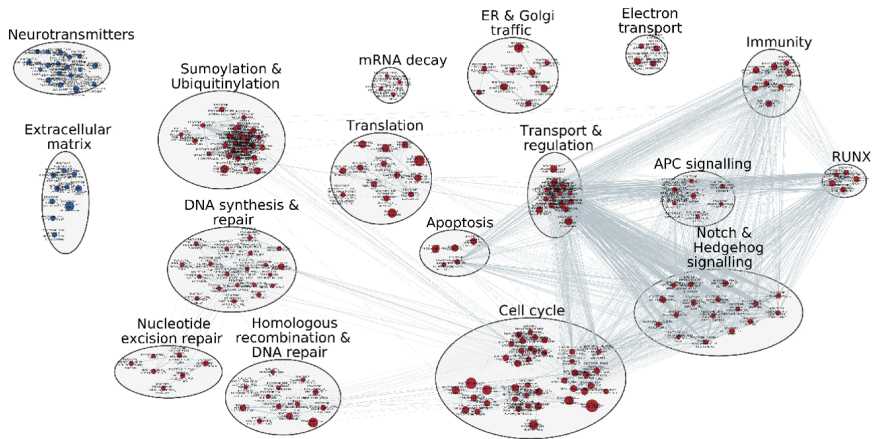


**Figure S9. Relation between tumor-infiltrating estimated immune cells abundances and immune biomarkers and treatment response.** Boxplots comparing **a. abundance** of immune cells and treatment response and **b. expression** of different immune-related genes (log2CPM) and signatures (GSVA scores). Immune cell infiltration was estimated with CIBERSORT using gene expression profiles. P-values were calculated using Wilcoxon rank sum test. Details on immune biomarkers and immune cell profiles can be found in the supplementary data file 2. NR: no responders, R: responders triangle shape represents the complete responders among the responder group.

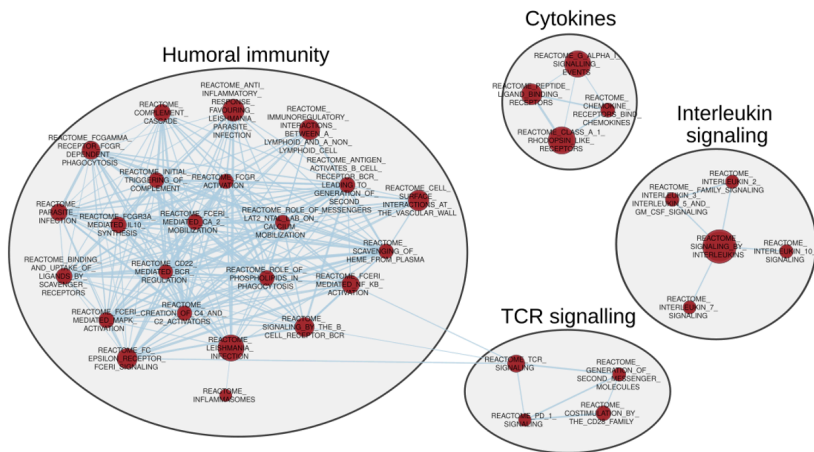


**Figure S10. Pathways of gene set enrichment analysis significantly related to ICI response** (adjusted  $P < 0.05$ , comparing 13 responders and 7 non-responders). Selected pathways are shown. The complete list of pathways with their adjusted p-value, normalized enrichment score and the included genes is provided in the Additional data file 3.

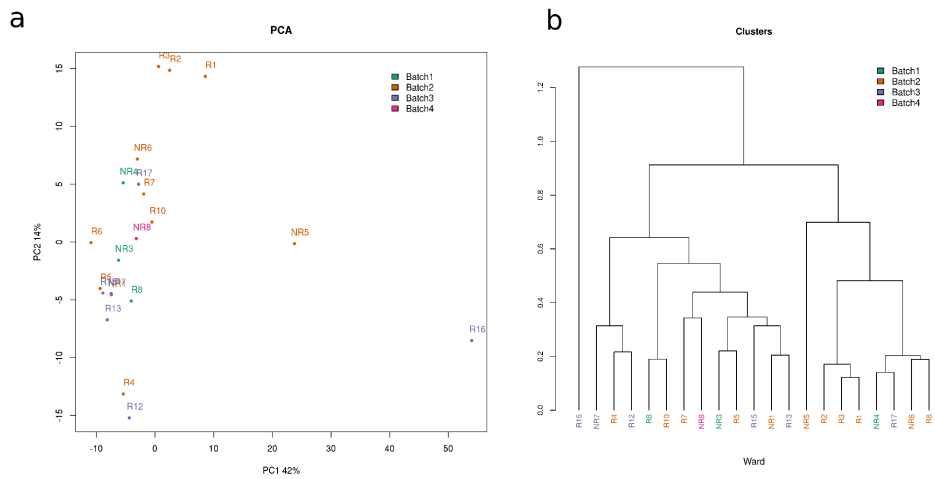
a



b

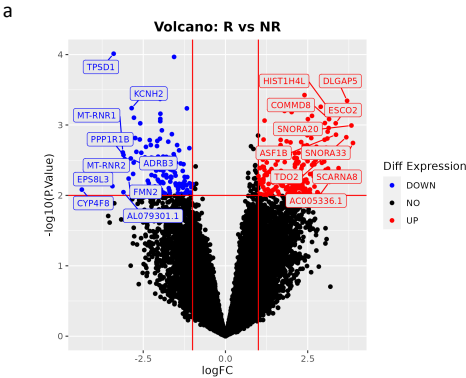


**Figure S11. Network of significantly enriched pathways.** Enrichment map of GSEA results (adjusted p-value < 0.05) comparing **a.** responders versus non-responders and **b.** complete responders versus partial-responders. Nodes represent genesets and edges represent the connectivity between genesets (combined metric Jaccard+Overlap >0.375). Red and blue represent positive and negative enrichment scores, respectively.

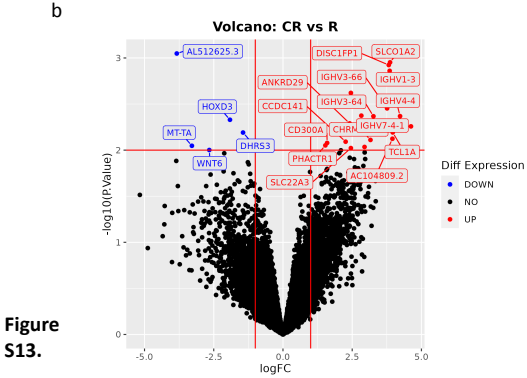


**Figure S12. Quality control of RNASeq data. a.** PCA using the top 500 most variable genes. **b.** Dendrogram using the top 500 most variable genes, 1-correlation distance and Ward2 linkage method. R16 was detected as an outlier. No batch effect was present.

a



b



**Figure S13.**

**Differentially expressed genes.** Volcano plots showing the results of the contrasts: **a.** responders vs non-responders. **b.** complete responders vs partial responders. The thresholds used in the plots are P.Value < 0.01 and |logFC|>1

## 5.2 Supplementary Material for Results 3.2

Supplementary figures and tables for the publication titled 'Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts', presented in section 3.2.

### **Full citation:**

Boll, L.M., Vázquez Montes de Oca, S., Camarena, M.E. et al. Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts. Nat Commun 16, 1213 (2025). <https://doi.org/10.1038/s41467-025-56462-0>.

**Sup. Table 1. Clinical and demographic information by ICI response groups (N=466).** Table of all patients with response status complete or partial response (R) or non-response (NR). Table generated with compareGroups R package using the default tests (t-test and ANOVA for continuous normal-distributed, X<sup>2</sup>-test for categorical variables, pairwise comparisons for more than 2 groups of continuous variables with Pearson test and for categorical with Mantel-Haenszel test, in both cases applying the Turkey method to adjust for multiple testing).

	NR N=303	R N=163	p.overall
<b>RECIST:</b>			<b>6.4e-102</b>
Complete Response	-	56 (34.4%)	
Partial Response	-	107 (65.6%)	
Progressive Disease	303 (100%)	-	
<b>Sex:</b>			<b>0.196</b>
Female	73 (24.1%)	30 (18.4%)	
Male	230 (75.9%)	133 (81.6%)	
<b>ECOG</b>	<b>0.80 (0.72)</b>	<b>0.66 (0.77)</b>	<b>0.060</b>
<b>ECOG ≥ 1:</b>			<b>0.006</b>
No	186 (66.0%)	75 (51.7%)	
Yes	96 (34.0%)	70 (48.3%)	
<b>Age</b>	<b>65.6 (10.6)</b>	<b>67.1 (10.3)</b>	<b>0.319</b>
<b>Race</b>			<b>0.504</b>
Asian	0 (0.00%)	1 (1.85%)	
Black or African American	4 (4.35%)	2 (3.70%)	
White	79 (85.9%)	49 (90.7%)	
<b>Smoker:</b>			<b>0.143</b>
No	86 (30.4%)	31 (23.0%)	
Yes	197 (69.6%)	104 (77.0%)	
<b>Alive:</b>			<b>2.2e-10</b>
No	92 (76.0%)	23 (29.5%)	
Yes	29 (24.0%)	55 (70.5%)	
<b>OS</b>	<b>119 (384)</b>	<b>269 (455)</b>	<b>4.7e-04</b>
<b>Liver metastasis:</b>			<b>0.008</b>
No	165 (66.8%)	79 (85.9%)	
Yes	82 (33.2%)	13 (14.1%)	
<b>ICI Drug:</b>			<b>0.043</b>
Anti-CTLA-4 + anti-PD-1/PD-L1	0 (0.00%)	1 (0.83%)	
Anti-PD-1/anti-PD-L1	10 (3.79%)	12 (9.92%)	
Atezolizumab	202 (76.5%)	88 (72.7%)	
Avelumab	2 (0.76%)	3 (2.48%)	
Durvalumab	4 (1.52%)	1 (0.83%)	
Nivolumab	5 (1.89%)	0 (0.00%)	
Pembrolizumab	41 (15.5%)	16 (13.2%)	



**Sup. Table 2. Abundance of different types of non-synonymous mutations.** WES data from the tumors of 318 patients was obtained using the same pipeline. The cohorts were IMvigor210, MIAO-2018, HdM-BLCA-1 and SNY-2017. R: responder; NR: non-responder.

<b>Mutation classification</b>	<b>Total (N=318)</b>	<b>R (N=89)</b>	<b>NR (N=144)</b>
<b>Frameshift deletion</b>	1373	673	34
<b>Frameshift insertion</b>	507	197	178
<b>In-frame deletion</b>	277	123	93
<b>In-frame insertion</b>	19	3	13
<b>Missense mutation</b>	55652	24423	18006
<b>Nonsense mutation</b>	5179	2225	1696
<b>Non-stop mutation</b>	85	40	28
<b>Splice site</b>	1102	428	414
<b>Translation start site</b>	81	35	27
<b>Total</b>	<b>64275</b>	<b>28147</b>	<b>20849</b>
Mean number of mutations	202	316	144

**Sup. Table 3. Most frequently mutated genes with the number of non-synonymous mutations combined for the four WES cohorts (IMvigor210, MIAO-2018, HdM-BLCA-1 and SNY-2017). Genes are sorted from most to fewest mutations found for the gene, only showing genes with more than 24 somatic mutations.**

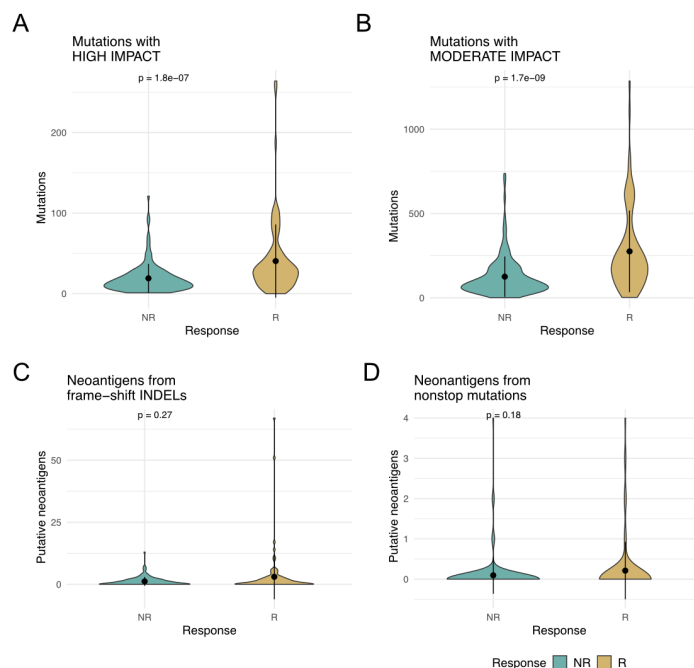
[illegible]

**Sup. Table 4. AUC values for different random forest models.** Baseline model corresponds to using only the variable TMB. Complete model is obtained with 17 variables, including mutation, gene expression and clinical variables. TMB + RNA model includes TMB and gene expression variables. The area under the curve (AUC) is the average of 1000 runs. Data is shown for the split-sample method and the internal bootstrapping method.

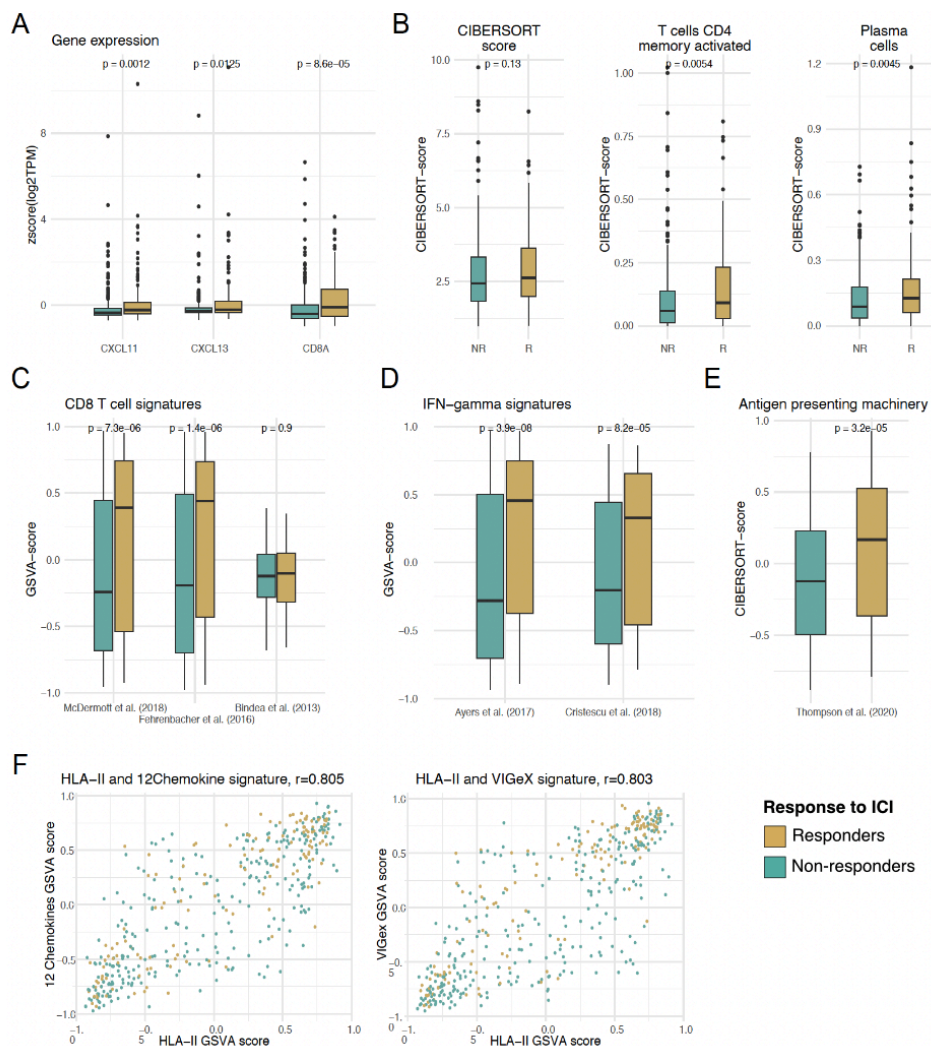
Model	Split-sample AUC	Bootstrap .632+ AUC
Baseline model, complete data	0.678	0.711
Baseline model, immune-infiltrated data	0.704	0.737
Baseline model, not-immune-infiltrated data	0.652	0.664
Complete model, complete data	0.761	0.789
Complete model, immune-infiltrated data	0.793	0.817
Complete model, not-immune-infiltrated data	0.639	0.646
TMB + RNA model, complete data	0.747	0.773
TMB + RNA model, immune-infiltrated data	0.769	0.792
TMB + RNA model, not-immune-infiltrated data	0.647	0.643

**Sup. Table 5. Random forest models to predict response to ICI in different bladder cancer subtypes.**  
The variables selected were TMB z-score, clinical variables (ECOG, liver metastasis) and RNA-Seq-derived variables.

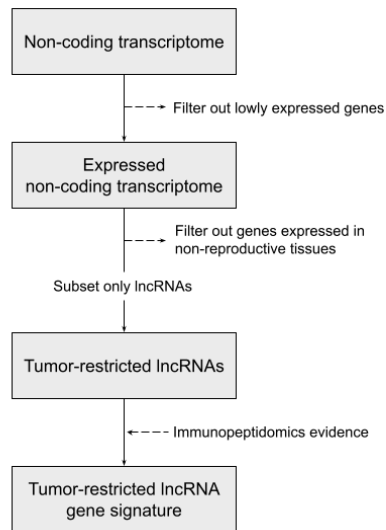
<b>Luminal Papillary</b> N=71; AUC (train/test) = 0.6544	<b>Luminal Infiltrated</b> N=138; AUC(train/test) = 0.8022	<b>Basal squamous</b> N=91; AUC (train/test) = 0.8223
Macrophages M1: 0.109 TMB_zscore: 0.106 PD1.zscore: 0.093 IFNg_Ayers.GSVA: 0.074 T cells CD8: 0.069 CCND1: 0.066 Stroma_EMT.GSVA: 0.063 PDL1.zscore: 0.059 APM_8.GSVA: 0.058 TGF_beta.GSVA: 0.049 T cells CD4 memory activated: 0.048 HLA-I.GSVA: 0.047 T_cell_inflamed.GSVA: 0.044 T cells regulatory (Tregs): 0.042 t.spec.lncRNA.GSVA: 0.038 ECOG_over0_Y: 0.017 Liver.Metastasis_Y: 0.017	TMB_zscore: 0.169 Macrophages M1: 0.13 IFNg_Ayers.GSVA: 0.086 T_cell_inflamed.GSVA: 0.066 t.spec.lncRNA.GSVA: 0.063 TGF_beta.GSVA: 0.062 Stroma_EMT.GSVA: 0.06 APM_8.GSVA: 0.058 HLA-I.GSVA: 0.05 CCND1: 0.048 PDL1.zscore: 0.046 PD1.zscore: 0.044 T cells CD4 memory activated: 0.042 T cells CD8: 0.037 T cells regulatory (Tregs): 0.022 ECOG_over0_Y: 0.009 Liver.Metastasis_Y: 0.007	TMB_zscore: 0.146 Macrophages M1: 0.085 PD1.zscore: 0.084 IFNg_Ayers.GSVA: 0.083 HLA-I.GSVA: 0.08 CCND1: 0.075 APM_8.GSVA: 0.064 T cells CD8: 0.063 T cells regulatory (Tregs): 0.059 PDL1.zscore: 0.053 t.spec.lncRNA.GSVA: 0.041 Stroma_EMT.GSVA: 0.04 TGF_beta.GSVA: 0.04 T_cell_inflamed.GSVA: 0.036 T cells CD4 memory activated: 0.032 Liver.Metastasis_Y: 0.011 ECOG_over0_Y: 0.008



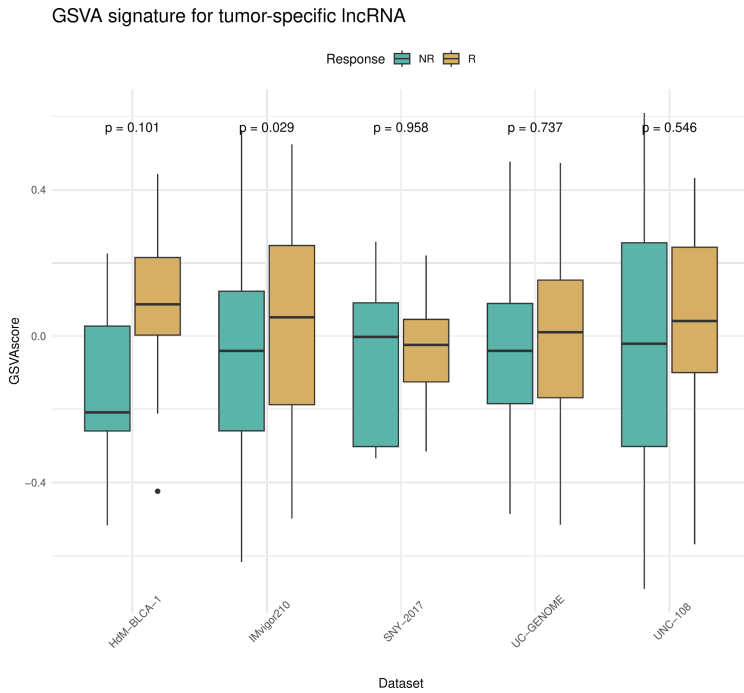
**Sup. Figure 1.** Extended list of mutation-based variables. **A.** Responders have a higher number of mutations annotated as HIGH impact (two-sided, two-sample Wilcoxon test  $p$ -value= $1.8e-079$ ). **B.** Similarly the number of MODERATE impact mutations is also significantly higher in responders than in non-responders (two-sided, two-sample Wilcoxon test). **C.** The number of putative neoantigens resulting from frameshift insertions and deletions (fsINDELs) is not significantly different between responders and non-responders (two-sided, two-sample Wilcoxon test). **D.** The number of putative neoantigens resulting from nonstop mutations was not found to be significantly associated with response (two-sided, two-sample Wilcoxon test). Mutation impact was obtained from ENSEMBL vep annotation (version 104). Putative binding peptides were predicted by applying a threshold of 500nM IC50 binding affinity in NetMHCpan 4.0. N=236 patients. R: responders (yellow), NR: non-responders (turquoise). All  $p$ -values are indicated in the according plots. Source data is provided as a Source Data file.



**Sup. Figure 2. Extended list of markers of immune activity in the tumor microenvironment. A.** Responders have higher gene expression of chemokines and CD8A compared to non-responders (two-sided, two-sample Wilcoxon test). **B.** Although responders do not show a higher general immune cell infiltration following the CIBERSORT absolute score, they have higher infiltration of cells related to an anti-tumor response by the immune system (two-sided, two-sample Wilcoxon test p-value=0.13). **C.** Responders are enriched in three out of four CD8 T cell signatures (two-sided, two-sample Wilcoxon test p-value=0.0054). **D.** Responders are enriched in IFN-gamma gene signatures (two-sided, two-sample Wilcoxon test). **E.** Responders are enriched in genes related to the antigen-presenting machinery (two-sided, two-sample Wilcoxon test). **F.** High correlation between signatures of immune activation and HLA-II expression (two-sided, two-sample Wilcoxon test). N=420 patients. R: responders (yellow), NR: non-responders (turquoise). All p-values are indicated in the according plots. Source data are provided as a Source Data file.

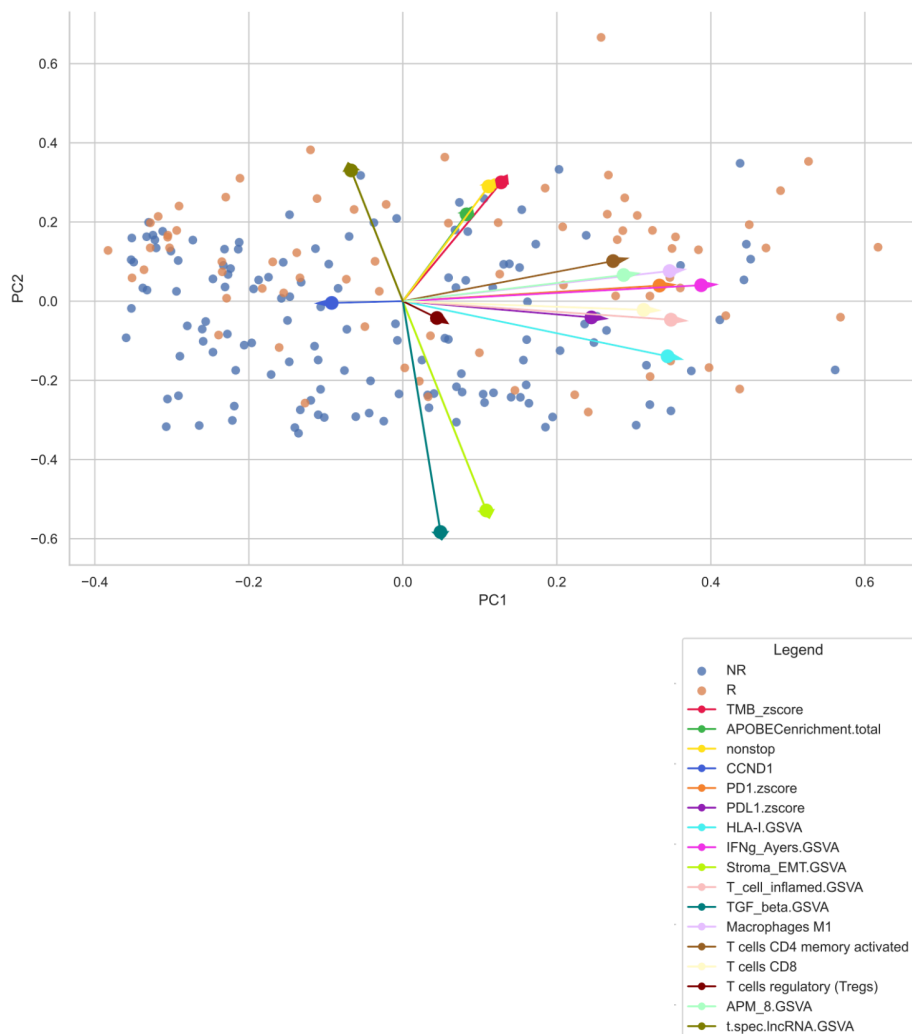


**Sup. Figure 3. Workflow overview of the identification of a tumor-restricted lncRNA gene signature.** Designed approach to detect tumor-restricted long non-coding RNAs and processed pseudogenes, collectively named lncRNAs. Cohorts were treated independently.

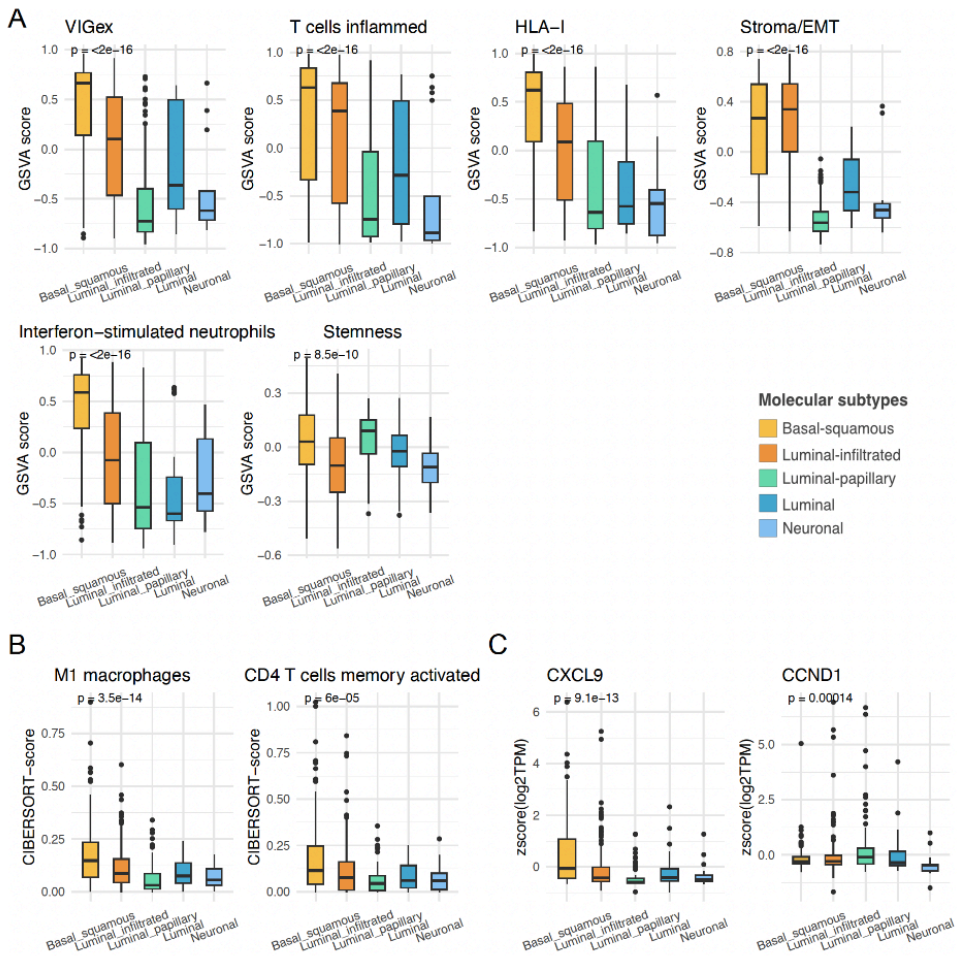


**Sup. Figure 4. Tumor-specific long non-coding RNAs gene signature.** GSVA signature score based on tumor-specific lncRNAs with immuno-peptidomics evidence per dataset (two-sided, two-sample Wilcoxon test: HdM-BLCA-1 p-value=0.101, IMvigor210 p-value=0.029, SNZ-2017 p-value=0.958, UC-GENOME p-value=0.737, UNC-108 p-value=0.546). N=420 patients. R: responders (yellow), NR: non-responders (turquoise). All p-values are indicated in the according plots. Source data are provided as a Source Data file.



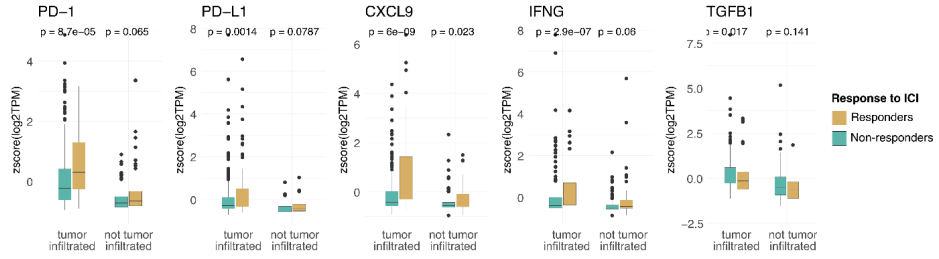


**Sup. Figure 5.** Principal component analysis of mutation and expression variables (N=205 patients). Patients, colored by response, are represented by the smaller points in the plot, while variables are represented as vectors. The orientation and length of the vectors denote the relationship between variables and observations. Variables closer together in the biplot indicate higher correlation.

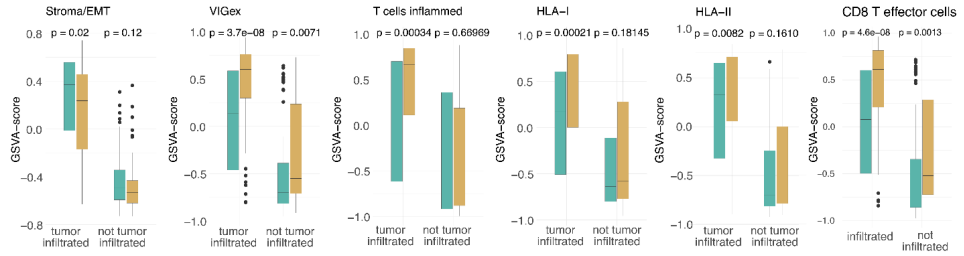


**Sup. Figure 6. Extended list of immune biomarkers in the five different TCGA subtypes. A.** Tumors of the basal-squamous subtype have the highest values of gene signatures related to immune activation as well as pathways of immune evasion such as Stroma/EMT (Kruskal-Wallis test). **B.** Basal-squamous is found to have the highest infiltration of CD8 T cells and M1 macrophages (Kruskal-Wallis test). **C.** Basal-squamous has the highest *CXCL9* expression. Luminal-papillary has the highest mean of *CCND1* expression (Kruskal-Wallis test). N=420 patients. Ba/Sq: Basal-squamous (yellow), LumInf: Luminal-infiltrated (orange), LumP: Luminal-papillary (green), Lum: Luminal (dark blue), NE: Neuronal (light blue). All p-values are indicated in the according plots. Source data are provided as a Source Data file.

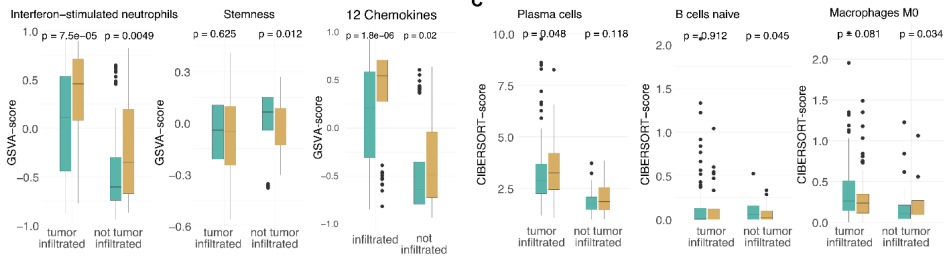
**A**



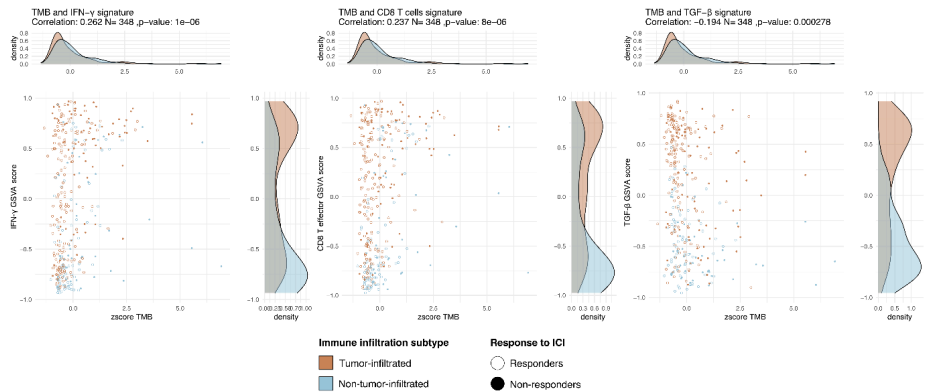
**B**



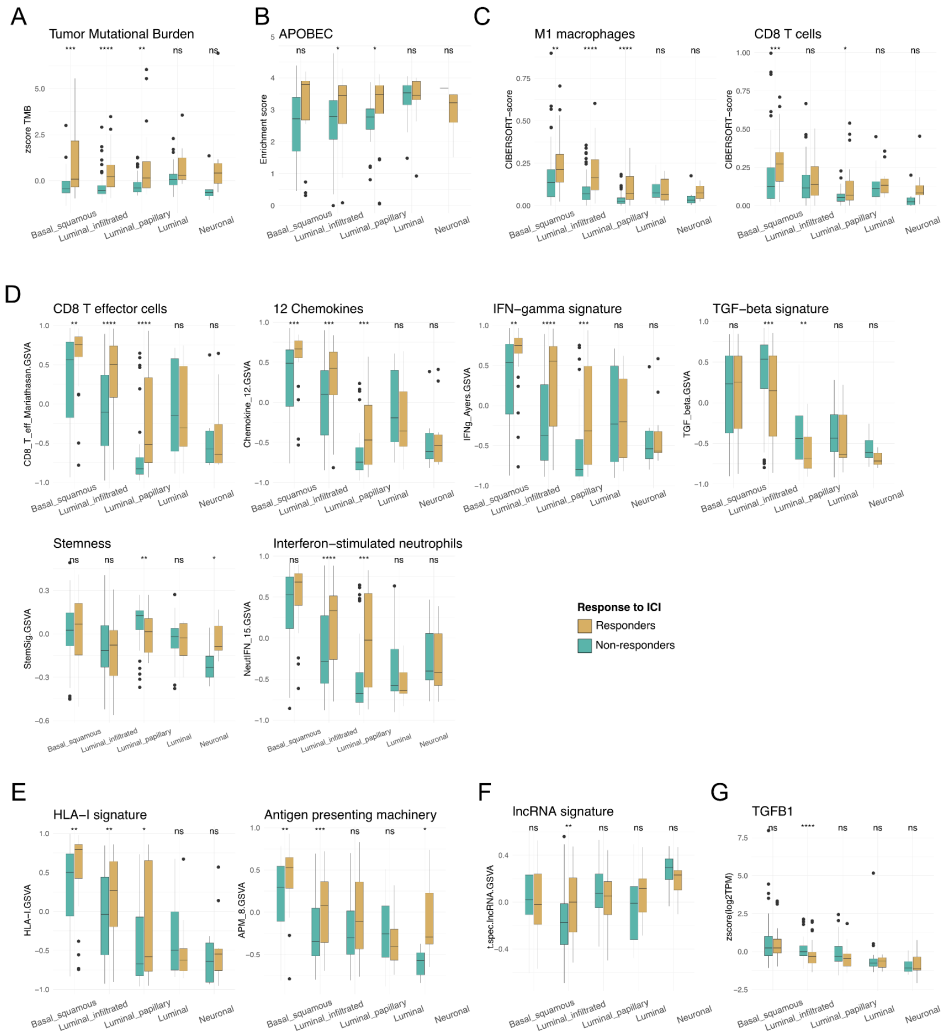
**C**



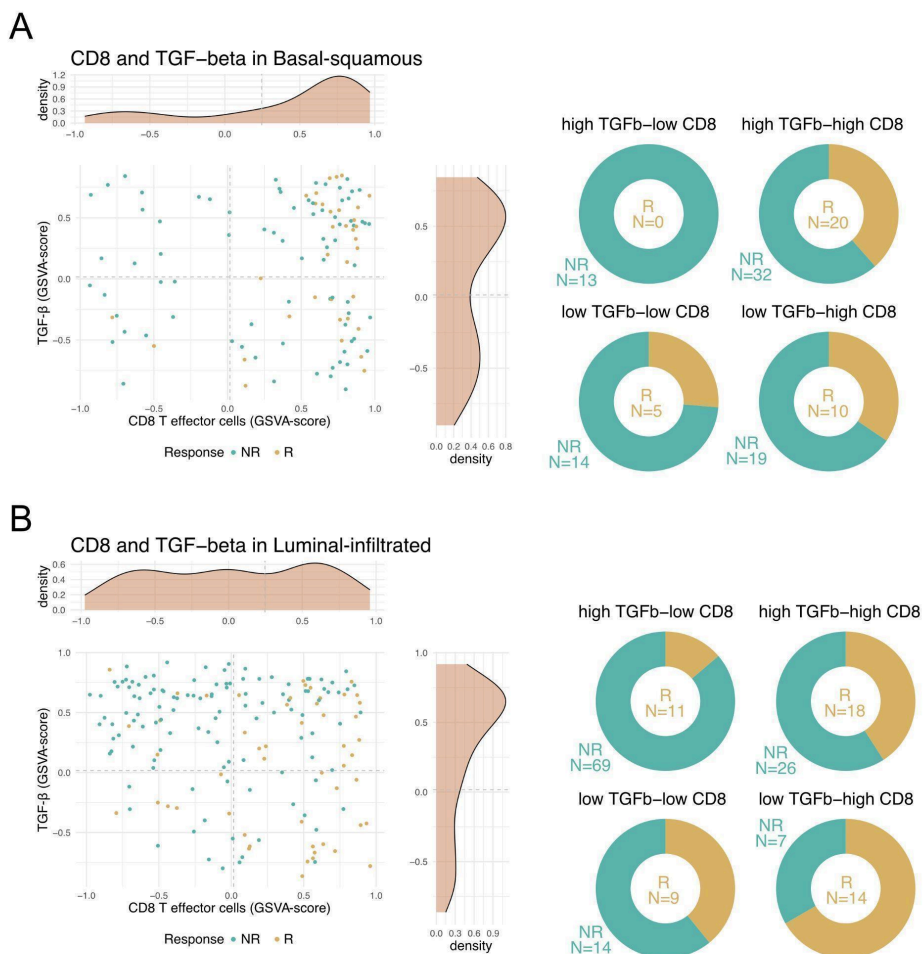
**D**



**Sup. Figure 7. Extended list of immune biomarkers in the immune-infiltrated and non-infiltrated subtypes.** **A.** Responders of the immune-infiltrated subtype have a higher mean expression of *TGFB1* than non-responders. No difference was observed for the not-infiltrated subtype (two-sided, two-sample Wilcoxon test). **B.** Data for gene signatures (two-sided, two-sample Wilcoxon test). **C.** Data for immune cell deconvolution analysis (two-sided, two-sample Wilcoxon test). A full list of the included genes and sources for the gene signatures is provided in supplementary file 2. **D.** Relationship between TMB and immune-related variables, in immune-infiltrated subtypes (red) and non-immune-infiltrated subtypes (blue), responders are marked with circles, non-responders with filled circles. N=420 patients. R: responders (yellow), NR: non-responders (turquoise). All p-values are indicated in the according plots. Source data are provided as a Source Data file.



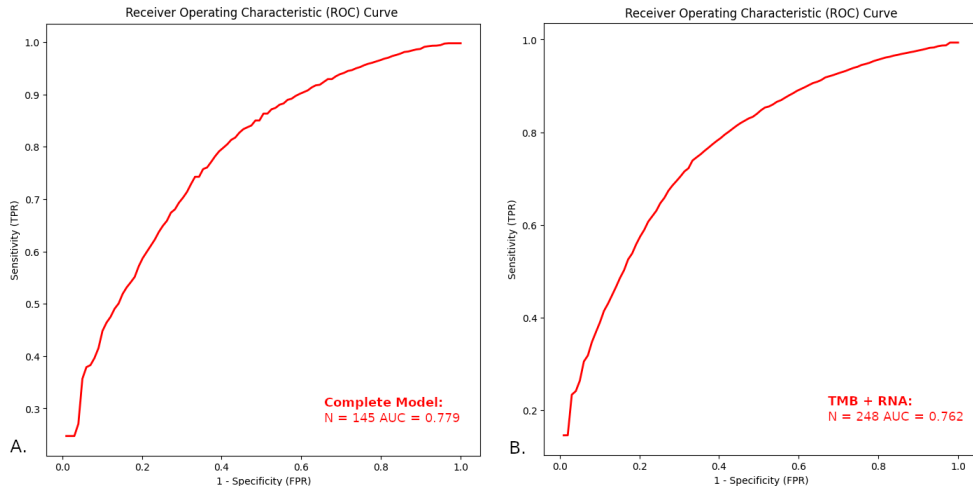
**Sup. Figure 8. Extended list of mutational and immune biomarkers in responders and non-responders by TCGA subtypes.** A. Relationship between TMB and subtype/response. B. Relationship between APOBEC enrichment and subtype/response. C. Relationship between immune cells and subtype/response. D. Relationship between immune activation and suppression signatures and subtype/response. E. Relationship between antigen presenting machinery (APM) pathway and subtype/response. F. Relationship between the tumor specific lncRNA signature and subtype/response. G. Relationship between TGFBI gene expression and subtype/response. N=420 patients. P-values obtained by two-sample Wilcoxon test ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ . R: responders (yellow), NR: non-responders (turquoise). All p-values are indicated in the according plots. Source data are provided as a Source Data file.



**Sup Figure 9. Relationship between CD8 T cell and TGF- gene expression in immune-infiltrated subtypes.** Relationship between CD8 T cell and TGF- $\beta$  gene expression by the immune-infiltrated TCGA subtypes basal-squamous (A) and luminal-infiltrated (B). Immune-infiltrated samples tend to have high CD8+ T cell abundance, and in many cases also high levels of the TGF- $\beta$  signature. Dashed line marks the optimal cutoff for each gene signature obtained by ROC (TGF- $\beta$ : 0.0163; CD8 t effector cells: 0.246). N=420 patients. R: responders (yellow), NR: non-responders (turquoise).



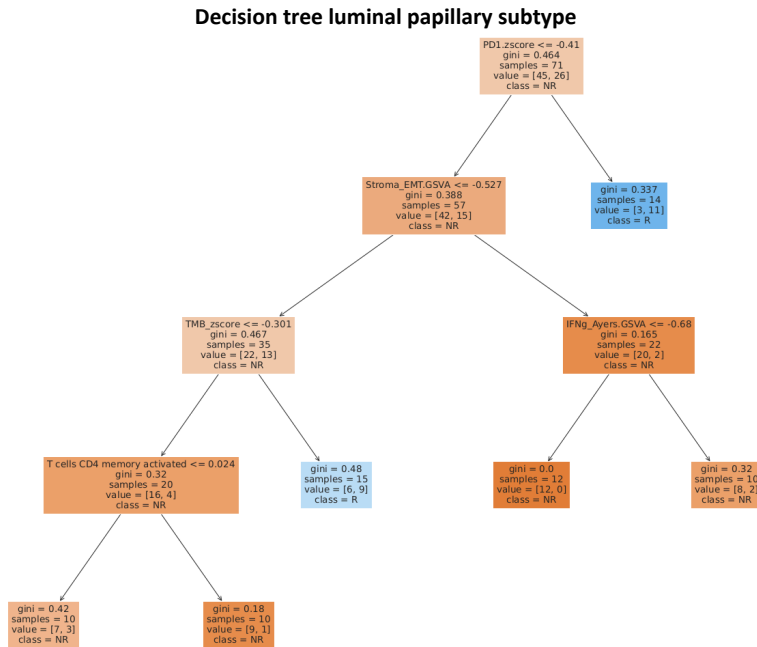
**Sup. Figure 10. Feature importance of all variables in the final model.** The data is shown for immune-infiltrated subtypes (N=137) and not-immune infiltrated subtypes (N=68). Bars are colored according to their association with therapy response (green: higher in responders; red: higher in non-responders). Source data are provided as a Source Data file.



**Sup. Figure 11. Performance of the prediction model when undersampling.** The models were built with an equal number of responder and non-responders to assess performance without group imbalance. **(A)** Shows results for the Complete model and **(B)** reports the results for the TMB + RNA Model (TMB z-score and RNA-Seq-derived variables).

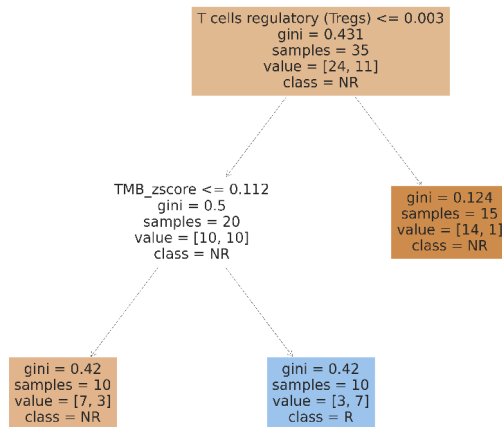


A.



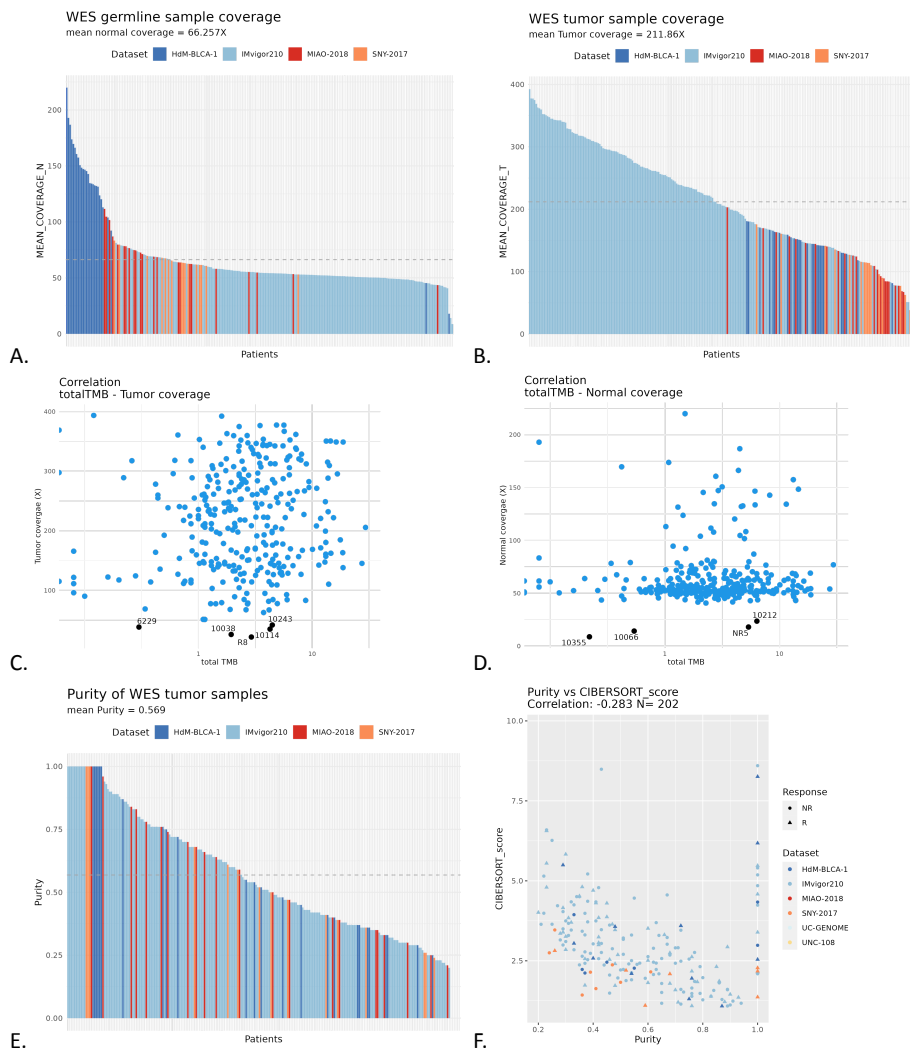
B.

**Decision tree luminal subtype**

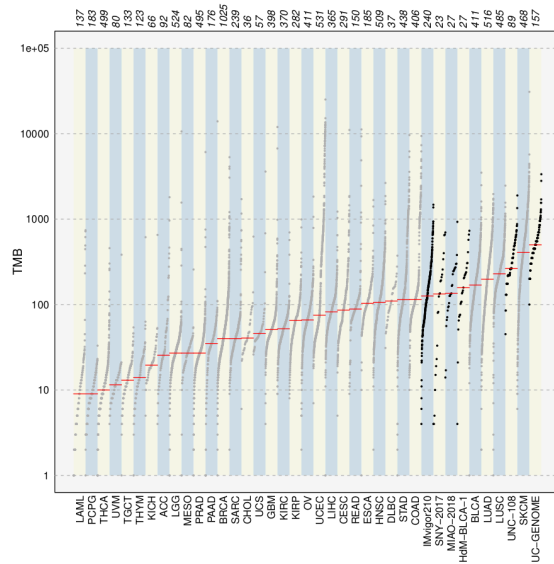


**Sup. Figure 12. Decision trees of non-immune-infiltrated subtypes.** Decision trees are created to better understand which features drive R/NR in our data stratifying by luminal-papillary (A) and luminal (B) TCGA Subtypes. The variables considered were TMB z-score and RNA-Seq-derived variables. The number of samples was 71 for luminal papillary and 35 for luminal. Number of leaves was set to 10.



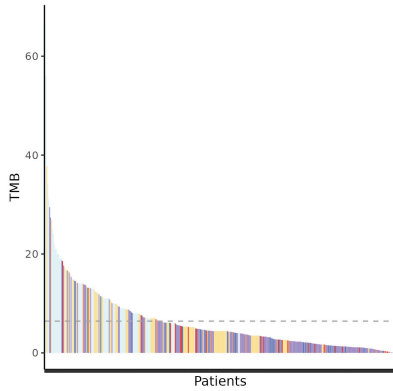


**Sup. Figure 14. Quality control of normal and tumor whole exome sequencing (WES) samples.** **A.** Coverage depth of germline WES samples. Samples from HdM-BLCA-1 were covered with a higher depth compared to the other three datasets. **B.** Coverage depth of WES from tumor samples. Samples of IMvigor210 have the highest coverage. **C.** Scatterplot showing the correlation between tumor mutational burden and tumor sample depth. Samples with a coverage > 20X are marked in black. As they don't fall into the 5% of samples with lowest mutation burden, they were not excluded in the analysis. **D.** Patients with a normal sample of low coverage don't show an unexpectedly high mutational burden and were therefore not excluded from the analysis. **E.** Purity estimates of tumor samples obtained from WES data using ASCAT. **F.** Correlation between WES tumor sample purity and deconvolution value obtained from CIBERSORT using tumor RNA-Seq data. Responders are marked with triangles, non-responders with circles.



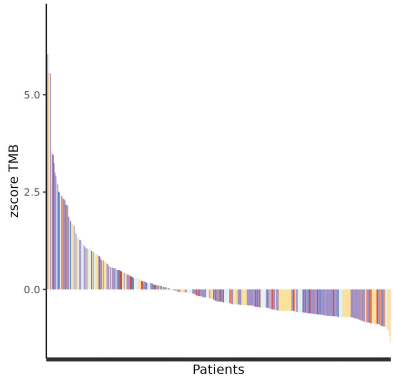
A.

TMB in the different datasets (mutants/Mb)  
mean TMB = 6.424 muts/Mb



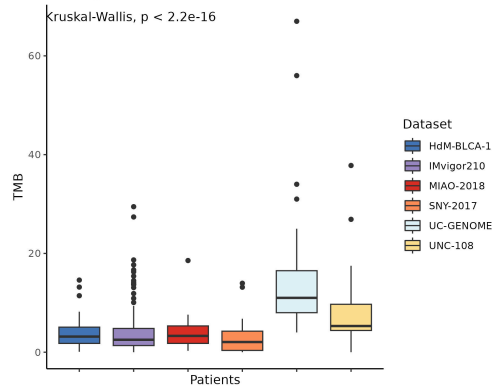
B.

Normalized TMB in the different datasets  
mean zscore TMB = 0.067

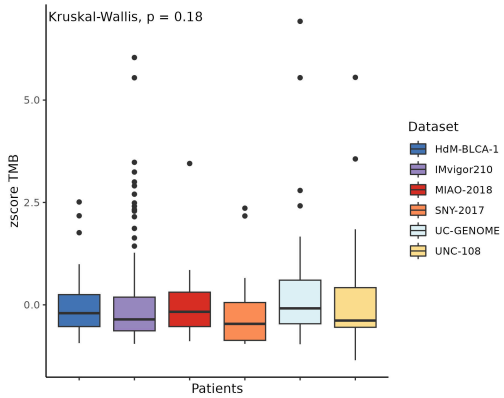


C.

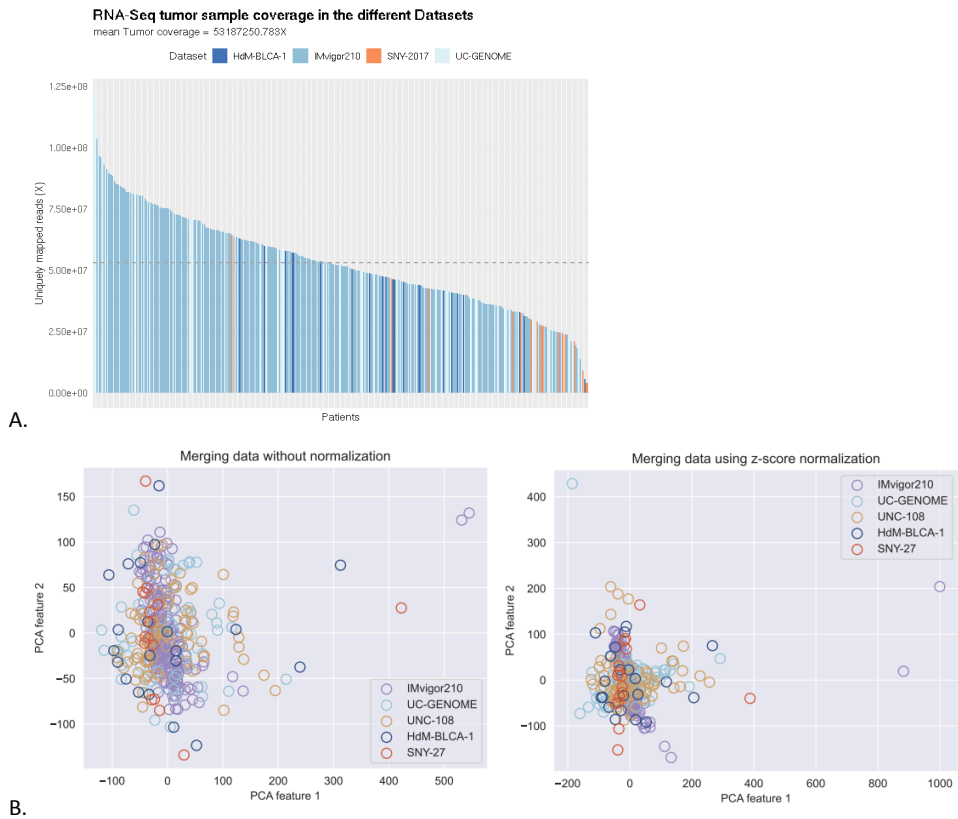
TMB by datasets



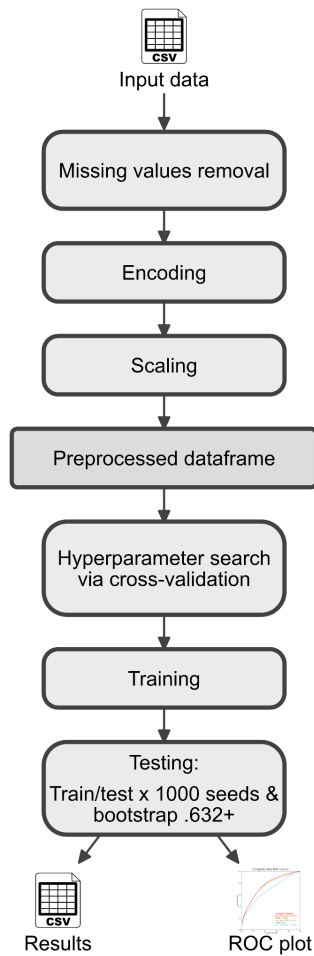
zscore TMB by datasets



**Sup. Figure 15. Normalization of TMB values.** **A.** Distribution of TMB in the different datasets compared to other TCGA cohorts. The four datasets with WES data (IMvigor210, SNY-2017, MIAO-2018 and HdM-BLCA-1) show a similar number of mutations as the TCGA bladder cancer cohort (BLCA). The two datasets with panel DNA (UC-GENOME and UNC-108) show overall higher TMB estimates. **B.** TMB values sorted descendingly and coloured by dataset (Kruskal-Wallis test). **C.** Sorted z-score TMB coloured by dataset (Kruskal-Wallis test). HdM-BLCA-1 (blue), IMvigor210 (purple), MIAO-2018 (red), SNY-2017 (orange), UC-GENOME (light grey), UNC-108 (yellow). Source data of the boxplots are provided as a Source Data file.



**Sup. Figure 16. Quality control and normalization of tumor RNA-Seq samples. A.** Uniquely mapped reads of all RNA-Seq samples, sorted descendingly and coloured by dataset. The horizontal line indicates the mean value. **B.** Batch effect of the five different datasets when conducting a PCA on the  $\log_2(\text{TPM}+1)$  matrices combining all RNA-Seq data. Normalizing the expression data using z score corrects for the batch effect. HdM-BLCA-1 (blue), IMvigor210 (purple), MIAO-2018 (red), SNY-2017 (orange), UC-GENOME (light grey), UNC-108 (yellow).



**Sup. Figure 17. Model flow chart.** The process of training and testing the model comprises multiple steps that start with data gathering followed by preprocessing steps and finally training and testing the model. A file containing all the results and a ROC plot are obtained.





## 5.3 Supplementary Material for Results 3.3

Supplementary material for the manuscript titled 'Stop-loss mutations in cancer', presented in section 3.3.

**Table S1: Number of samples, missense, and stop-loss mutations in TCGA cancer types.** The data corresponds to the TCGA MC3 project, sorted by the proportion of stop-loss mutations.

Cancer Type	Acronym	Patients	Missense Mutations	Stop-loss Mutations	Proportion of Stop-loss
Testicular Germ Cell Tumors	TGCT	134	1710	9	0.52
Acute Myeloid Leukemia	LAML	140	4137	18	0.43
Kidney Chromophobe	KICH	66	1851	4	0.22
Uterine Carcinosarcoma	UCS	57	6098	12	0.20
Pheochromocytoma and Paraganglioma	PCPG	184	1600	3	0.19
Kidney Renal Papillary Cell Carcinoma	KIRP	282	16655	31	0.19
Thyroid Carcinoma	THCA	500	6656	12	0.18
Kidney Renal Clear Cell Carcinoma	KIRC	370	16250	29	0.18
Thymoma	THYM	123	2248	4	0.18
Bladder Urothelial Carcinoma	BLCA	411	85142	150	0.18
Liver Hepatocellular Carcinoma	LIHC	365	30935	49	0.16
Adrenocortical Carcinoma	ACC	92	6496	10	0.15
Cervical Squamous Cell Carcinoma	CESC	291	49790	74	0.15
Mesothelioma	MESO	82	2060	3	0.15
Ovarian Serous Cystadenocarcinoma	OV	411	29641	42	0.14
Esophageal Carcinoma	ESCA	185	21846	30	0.14
Lung Squamous Cell Carcinoma	LUSC	485	114119	153	0.13
Lung Adenocarcinoma	LUAD	517	122329	164	0.13
Breast Invasive Carcinoma	BRCA	1026	66867	89	0.13
Cholangiocarcinoma	CHOL	36	1643	2	0.12
Diffuse Large B-cell Lymphoma	DLBC	37	4171	5	0.12
Rectum Adenocarcinoma	READ	150	38487	45	0.12
Head and Neck Squamous Cell Carcinoma	HNSC	509	64476	74	0.11
Sarcoma	SARC	239	15512	17	0.11
Stomach Adenocarcinoma	STAD	439	116742	124	0.11
Glioblastoma Multiforme	GBM	400	40323	38	0.09
Endometrial Carcinoma	UCEC	531	459558	420	0.09
Colon Adenocarcinoma	COAD	406	137248	113	0.08
Brain Lower Grade Glioma	LGG	525	24779	18	0.07
Prostate Adenocarcinoma	PRAD	498	18539	13	0.07
Skin Cutaneous Melanoma	SKCM	468	296949	164	0.06
Pancreatic Adenocarcinoma	PAAD	178	18364	10	0.05
Uveal Melanoma	UVM	80	1186	0	0.00

**Table S2: Number of clonal and subclonal mutations. 58.94% of the stop-loss mutations and 62.48% of the missense mutations were clonal (chi-square p-value=0.7). We only considered missense mutations from TCGA patients with stop-loss mutations.**

	<b>Clonal</b>	<b>Subclonal</b>	<b>Total</b>
<b>Missense mutations</b>	1,147,171	688,885	1,836,056
<b>Stop-loss mutations</b>	1,141	795	1,936

**Table S3: Number of missense and stop-loss mutations in different groups of genes.** Significance is marked with \* accounting for multiple testing using the Bonferroni methods (alpha=0.01)

	Missense (N = 1,826,158)	Stop-loss (N = 2,066)	X <sup>2</sup> p-value
Cancer gene list (N = 77)	13,293 (0.73%)	26 (1.26%)	0.0045 *
COSMIC oncogenes (N = 319)	25,018 (1.36%)	44 (2.12%)	0.00297 *
COSMIC tumor suppressor genes N=320	29,628 (1.60%)	40 (1.93%)	0.25

**Table S4: Gene lists of cancer-related genes.** The mutations in these genes are indicated in Table S3.

Cancer-related genes (n = 77)	
DNA-damage repair	ATM, ATRX, BLM, BRCA1, BRCA2, BRIP1, CHEK2, DDB2, ERCC2, ERCC3, ERCC4, ERCC5, FANCA, FANCD2, FANCE, FANCF, FANCG, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS1, PMS2, PRKDC, WRN, XPA, FANCC, XPC
Chromatin-modifying	ARID1A, ARID2, ASXL1, ATRX, AURKB, BRCA1, BRCA2, CHEK2, CREBBP, EP300, KDM6A, MECOM, NPM1, RAD21, SMC3, STAG2
Chromatin-regulating	KDM6A, KMT2A, KMT2B, KMT2C, KMT2D, EP300, CREBBP, ARID1A, CDH6, ASXL1, ASXL2, KANSL1
HLA	HLA-F, HLA-G, HLA-A, HLA-E, HLA-C, HLA-B, HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-DQB2, HLA-DOB, HLA-DMB, HLA-DMA, HLA-DOA, HLA-DPA1, HLA-DPB1
COSMIC gene lists	
Oncogenes (n = 319)	TNFRSF17, CYSLTR2, BCL2L12, ARHGAP5, ZNF521, TRIM27, TRIM24, TMSB4X, STAT5B, SH3GL1, SETDB1, PDGFRB, PABPC1, NKX2-1, NFATC2, MALAT1, KNSTRN, HOXD13, HOXD11, HOXC13, HOXC11, HOXA13, CTNND2, CTNNA2, CHST11, CDKN1A, WWTR1, U2AF1, TRRAP, TAF15, STAT6, STAT3, SRSF3, SRSF2, RUNX1, RSPO3, PLAG1, PDGFB, P2RY8, OLIG2, NUTM1, NUP98, NTRK2, NR4A3, NFKB2, MUC16, MTCPI, MRTFA, MECOM, MAML2, MALT1, MACC1, KDM5A, KCNJ5, KAT6A, IKZF3, HOXA9, HMGA2, HMGA1, FSTL3, FOXR1, FOXP1, FOXO4, FOXO3, FOXO1, FCRL4, DGCR8, CSF1R, CREB1, CDH17, CCND3, BIRC6, ZEB1, USP6, TSHR, TLX3, TLX1, TFEB, TCF3, TBX3, TAL2, TAL1, STIL, SSX4, SSX2, SSX1, SOX2, SND1, SIX2, SGK1, RARA, POLQ, NSD3, NSD2, MYCN, MYCL, MUC4, MSI2, MTF, MDM4, MDM2, MAFB, LYL1, LMO2, LMO1, KAT7, IRF4, GRM3, GLI1, FLI1, ETV5, ETV4, ETV1, CD74, CD28, CCR7, CCR4, CBLC, BRD3, BCL6, BCL3, BCL2, ARAF, AKT3, AKT2, AFDN, ABL2, A1CF, WT1, TNC, TEC, SYK, SRC, SKI, SET, REL, MYB, MN1, MAF, LYN, LPP, JUN, HLF, FEV, DEK, AR, ALK, BTK, CBL, CIC, ERG, FES, KDR, KIT, LCK, MET, MPL, MYC, QKI, RET, SMO, WAS, ABL1, AFF3, AFF4, AKT1, ARNT, ATF1, BCL9, BRAF, BRD4, CALR, CDK4, CDK6, CHD4, CUX1, DAXX, DDB2, DDR2, DDX5, DDX6, EGFR, ELF4, ELK4, ESR1, EZH2, FLT3, FLT4, GNAQ, GNAS, GPC3, HEY1, HIP1, HRAS, IDH1, IDH2, IL7R, IRS4, JAK1, JAK2, JAK3, KLF4, KRAS, LEF1, MTOR, NPM1, NRAS, PAX3, PAX5, PBX1, PIM1, PTK6, RAC1, RAF1, RHOA, ROS1, SIX1, TERT, TET1, TFE3, TP53, TP63, UBR5, USP8, XPO1, ACKR3, ACVR1, BCL9L, BIRC3, CCND1, CCND2, CCNE1, CD79A, CD79B, CRLF2, CRTC1, CSF3R, CXCR4, DDIT3, EPAS1, ERBB2, ERBB3, ERBB4, EWSR1, FGFR1, FGFR2, FGFR3, FGFR4, FOXA1, FOXL2, FUBP1, GATA1, GATA2, GATA3, GNA11, H3F3A, H3F3B, HIF1A, IKBKB, IL6ST, KDM6A, KMT2A, KMT2D, MAPK1, MYD88, MYOD1, NCOA2, NT5C2, NTRK1, NTRK3, PLCG1, PPM1D, PREX2, PSIP1, RAD21, SALL4, SF3B1, SUZ12, TCL1A, ATP1A1, BCL11A, BCL11B, BCORL1, BMPR1A, CARD11, CREBBP, CTNNB1, FCGR2B, HOXA11, MAP2K1, MAP2K2, MAP2K4, MAP3K1, MLLT10, NFE2L2, NOTCH1, NOTCH2, PDGFRA, PIK3CA, PIK3CB, POU5F1, PRDM16, PRKACA, PTPN11, RECQL4, SETBP1, TCF7L2, CACNA1D, CREB3L2, MAP3K13, POU2AF1, PRKARIA, RUNX1T1, TBL1XR1, APOBEC3B, HIST1H3B, PDCD1LG2, RAP1GDS1, HNRNPA2B1

## COSMIC gene lists

### Tumor-suppressor genes (n = 320)

*FH, ARHGEF10L, TNFRSF14, LEPROTL1, ARHGEF12, ARHGEF10, ARHGAP35, TMEM127, SMARCE1, SMARCD1, IGF2BP2, CNTNAP2, ZNF331, TRIM33, TRIM24, TENT5C, STAT5B, SETD1B, SDHAF2, RAD51B, PABPC1, NKX2-1, MALAT1, LARP4B, CTNNA1, CDKN1A, ACVR1B, ZNRF3, ZMYM3, YWHAE, USP44, STK11, STAG1, SOX21, SMC1A, SIRPA, SFRP4, RUNX1, RSPO2, RPL22, RPL10, ROBO2, RFWF3, RAD17, PTPRK, PTPRD, PTPRC, PTPN6, PRDM2, PATZ1, PALB2, NTHL1, NFKB2, NCOA4, N4BP2, MUTYH, MRTFA, LRIG3, KAT6B, IKZF3, HOXA9, FOXO4, FOXO3, FOXO1, FBLN2, FANCG, FANCF, FANCE, FANCC, FANCA, CSMD3, CPEB3, CDH10, CD274, CASP9, CASP3, BAZ1A, ASXL2, WNK2, TSC2, TSC1, TPM3, TCF3, TBX3, SUFU, SDHD, SDHC, SDHB, SBDS, RMI2, PRF1, POLQ, POLG, NRG1, MYH9, MLF1, MGMT, KLF6, IRF4, GPC5, FLCN, FEN1, EXT2, EXT1, ELF3, CUL3, CHD2, CDX2, CCNC, CBLC, BTG2, BTG1, XPC, XPA, WT1, WRN, PML, NBN, ID3, EED, BAX, APC, ATM, ATR, B2M, BLM, BTK, CBL, CIC, ELL, FAS, FES, FUS, MAX, NF1, NF2, QKI, RB1, VHL, ABI1, ARNT, ATRX, BAP1, BCOR, CARS, CFBF, CBLB, CDH1, CLTC, CNBP, CTCF, CUX1, CYLD, DAXX, DDB2, DNM2, EBF1, ELF4, ESR1, ETV6, EZH2, FAT1, FAT4, FHIT, GPC3, IRS4, JAK1, KLF4, KNL1, LEF1, MEN1, MLH1, MSH2, MSH6, NAB2, PAX5, PER1, PHF6, PMS2, POLE, POT1, PTEN, PTK6, RHOA, RHOH, RPL5, SDHA, SFPQ, SPEN, SPOP, TERT, TET1, TET2, TP53, TP63, WIF1, AMER1, ARID2, ASXL1, AXIN1, AXIN2, BARD1, BCL10, BCL9L, BIRC3, BRCA1, BRCA2, BRIP1, BUB1B, CASP8, CCDC6, CDC73, CDH11, CDK12, CEBPA, CHEK2, CIITA, CNOT3, DDX10, DDX3X, EIF3E, EP300, EPAS1, EPS15, ERBB4, ERCC2, ERCC3, ERCC4, ERCC5, ETNK1, FBXW7, FOXL2, GATA1, GATA3, HNF1A, IKZF1, KDM5C, KDM6A, KEAP1, KMT2C, KMT2D, LATS1, LATS2, LRP1B, LZTR1, MED12, NCOR1, NCOR2, NDRG1, NTRK1, PBRM1, POLD1, PPARG, PPP6C, PRDM1, PTCH1, PTPRB, PTPRT, RAD21, RBM10, RNF43, SETD2, SH2B3, SMAD2, SMAD3, SMAD4, SOCS1, STAG2, SUZ12, TRAF7, ZFHX3, ZRSR2, ACVR2A, ARID1A, ARID1B, ATP1A1, ATP2B3, BCL11B, BCORL1, BMP1A, CAMTA1, CDKN1B, CDKN2A, CDKN2C, CLTCL1, CREBBP, DICER1, DNMT3A, DROSHA, FANCD2, FBXO11, GRIN2A, HOXA11, MAP2K4, MAP3K1, NFE2L2, NFKBIE, NOTCH1, NOTCH2, PHOX2B, PIK3R1, PTPN13, RANBP2, RECQL4, TGFB2, ZBTB16, CBFA2T3, CREB3L1, MAP3K13, PPP2R1A, PRKARIA, RUNX1T1, SLC34A2, SMARCA4, SMARCB1, TBL1XR1, TNFAIP3, APOBEC3B, ARHGAP26, CCNB1IP1*

**Table S5: Statistics of Hydrophobicity and isoelectric point distribution in stop-loss extensions, canonical sequences, and intronic ORFs.** P-values are obtained from the Wilcoxon test comparing stop-loss extensions to the other two groups.

Length	Min	Q1	Median	Mean	Q3	Max	
Stop-loss extension	1.00	8.00	17.00	26.45	36.00	221.00	

Hydrophobicity	Min	Q1	Median	Mean	Q3	Max	p-value
Stop-loss extension	-4.50	-0.83	-0.16	-0.18	0.40	4.50	
Canonical	-3.43	-0.59	-0.35	-0.33	-0.09	2.11	2.2e-16
iORF	-4.50	-0.47	0.07	0.09	0.64	4.50	9.4e-16

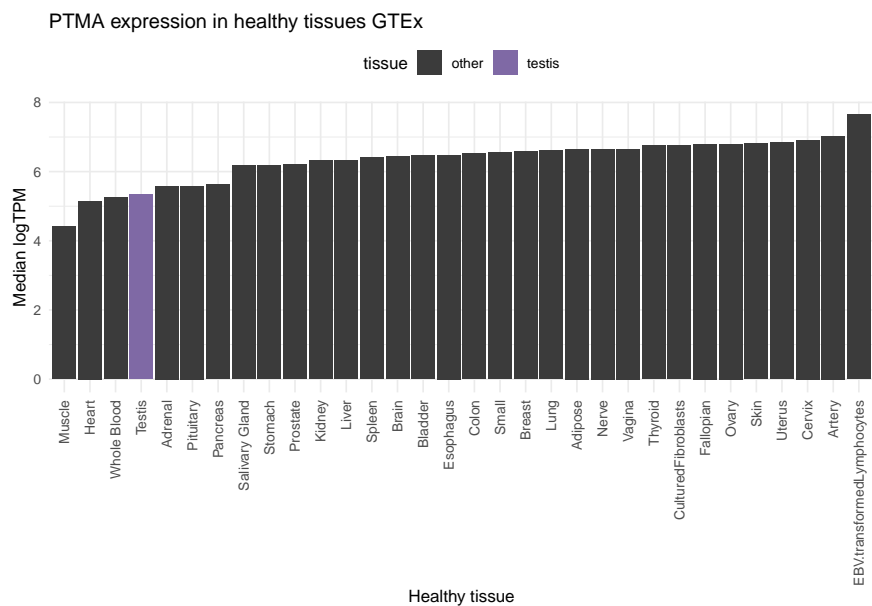
Isoelectric point	Min	Q1	Median	Mean	Q3	Max	p-value
Stop-loss extension	3.33	6.10	7.93	7.97	10.04	13.10	
Canonical	3.13	6.10	7.34	7.42	8.76	13.80	2.2e-16
iORF	2.18	6.20	8.52	8.16	10.13	14.00	0.006

<b>TAG</b>	*	<b>TAA</b>	*	<b>TGA</b>	*
TCG	Ser (S)	TCA	Ser (S)	GGA	Gly (G)
TAC	Tyr (T)	TAC	Tyr (T)	TCA	Ser (S)
TAT	Tyr (T)	TAT	Tyr (T)	TGC	Cys (C)
TGG	Trp (W)	TTA	Leu (L)	TTA	Leu (L)
TTG	Leu (L)	CAA	Gln (Q)	TGG	Trp (W)
CAG	Gln (Q)	GAA	Glu (E)	TGT	Trp (W)
GAG	Glu (E)	AAA	Lys (K)	AGA	Arg (R)
AAG	Lys (K)	TAG	*	CGA	Arg (R)
TAA	*	TGA	*	TAA	*

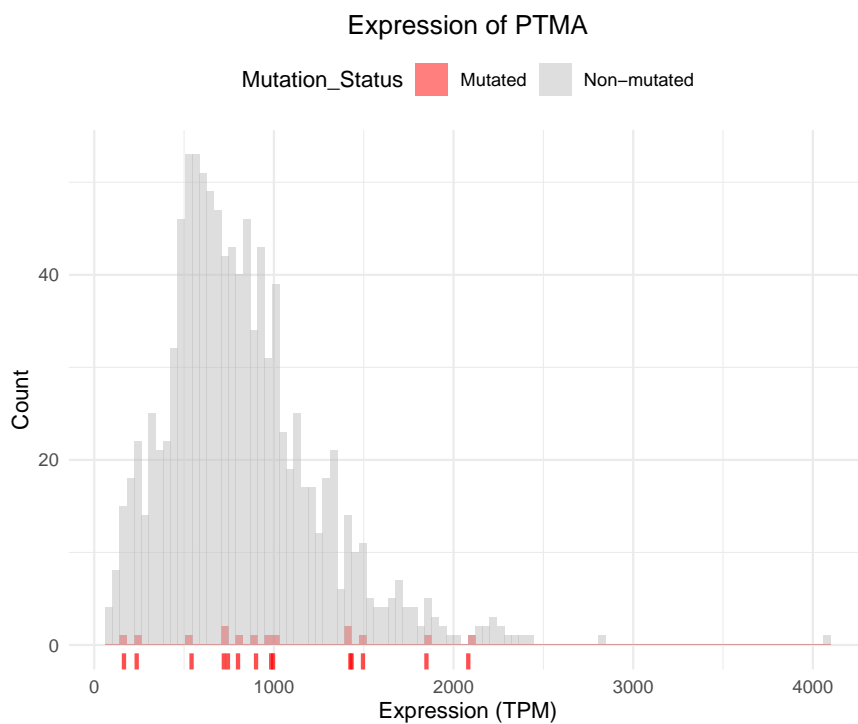
hydrophobic	polar	basic	neutral	acidic
-------------	-------	-------	---------	--------

**Figure S1: Possible nucleotide substitutions in the stop codons and the resulting codons.** The majority of stop codon mutations extend the coding sequence. Mutated codons with the encoded amino acid and its chemical properties. Variants that encode another stop codon (\*) or the same amino acid as a previous mutation in the list are marked in light grey. For each stop codon, 30% of the mutations result in a polar amino acid.



**Figure S2: Expression of the PTMA gene in different body tissues.** The median logTPM value in GTEx is shown.





**Figure S3: Expression of PTMA in tumors with the stop-loss mutation and tumors without this mutation.** Patients with a stop-loss mutation in PTMA are marked red. The expression in mutated samples is not significantly different from the distribution of expression values (two-sided t-test p-value=0.19).



## ANNEX

### 5.4 Academic Engagement

- **Research Stay at Vall d’Hebron Institute of Oncology (VHIO),** Barcelona, July 2024.
- **Committee Member of the International Society of Computational Biology (iSCB) Student Council,** 2024



### 5.5 Science Divuligation Talks

- **European Researcher’s Night 2021,** Barcelona, 24<sup>th</sup> of September 2021.
- **Dia Internacional de la Dona i la Nena a la Ciència 2021,** Orlot, 7<sup>th</sup> of February 2022.
- **OpenPRBB Day 2023,** Barcelona, 7<sup>th</sup> of October 2023.
- **Dia Internacional de la Dona i la Nena a la Ciència 2024,** Mataró, 6<sup>th</sup> of February 2024.
- **"Research in 4 minutes" (Rin4min) finals** by Universidad Pompeu Fabra, Barcelona, t<sup>th</sup> of June 2024<sup>1</sup>.
- **European Researcher’s Night 2024,** Barcelona, 27<sup>th</sup> of September 2024.

---

<sup>1</sup>On YouTube "Final Rin4min 2024" by Pompeu Fabra University - Barcelona (<https://www.youtube.com/watch?v=X-brz0mZ5mY&t=590s>)

## 5.6 Oral Presentations

- **PRBB Computational Genomics Seminar**, Barcelona, 4<sup>th</sup> of April 2024.
- **PRBB Computational Genomics Symposium**, Sant Feliu de Guíxols Spain, 28<sup>th</sup> of April 2025.
- **European Association for Cancer Research (EARC) Conference**, Barcelona, 13<sup>th</sup> of May 2025.

## 5.7 Poster Presentations

- **ACGT Computational Genomics workshops** by CRG and IMIM, Sant Feliu de Guíxols Spain, 13<sup>th</sup> of April 2023.
- **Immunity and Cancer Immunotherapy** by Institute Curie, Marseille, 21<sup>th</sup> of June 2023.
- **EMBL Cancer Genomics Conference**, Heidelberg, 16<sup>th</sup> of November 2023. Shown on the following page.
- **VHIO Computational Oncology Master Thesis Award**, Barcelona, 21<sup>st</sup> of February 2024.





## Colophon

This thesis was typeset with  $\text{\LaTeX}$  2 $_{\epsilon}$ . It uses the *Clean Thesis* style developed by Ricardo Langner.

