©2018

Anshuman Panda

ALL RIGHTS RESERVED

BIOMARKERS OF RESPONSE TO IMMUNE CHECKPOINT THERAPY

By

ANSHUMAN PANDA

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

and

Graduate Program in Quantitative Biomedicine

Written under the direction of

Prof. Gyan Bhanot

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

JANUARY, 2018

ABSTRACT OF THE DISSERTATION

Biomarkers of Response to Immune Checkpoint Therapy

By ANSHUMAN PANDA

Dissertation Director:

Prof. Gyan Bhanot

Immune checkpoint therapy leads to durable objective response in a subset of patients with metastatic cancer in many different cancer types, but the mechanisms and biomarkers of response to immune checkpoint therapy is not fully understood. Through pan-cancer analysis of the Cancer Genome Atlas, we have so far identified 2 types of predictors of immune activation and checkpoint pathway upregulation in a total of 10 solid cancer types (skin melanoma, lung adenocarcinoma, colon, endometrial, gastric, cervical, ER+ HER2− breast, bladder, clear-cell kidney, and head-neck squamous-cell cancer): (1) hyper-mutation in tumor due to DNA proofreading defect (Chapter 2) or some other etiologies (Chapter 3), and (2) expression of exogenous (Chapter 4) or endogenous (Chapter 5) viral RNA in tumor. We also validated these predictors as biomarkers of response to immune checkpoint therapy in some of the above 10 cancer types where we had access to good quality data, using published datasets of patients treated with immune checkpoint therapy for retrospective validation, and de-identified data of patients treated with immune checkpoint therapy at the Rutgers Cancer Institute of New Jersey and the Vanderbilt Ingram Cancer Center for prospective validation.

Acknowledgments

Published work

The work described in Chapter 2, 3, 4 are published in the Journal of Clinical Investigation, JCO Precision Oncology, and Journal of the National Cancer Institute respectively (as referenced below), and the work described in Chapter 5 is currently being prepared for publication.

Chapter 2: Mehnert JM, Panda A, Zhong H, et al. J Clin Invest. 2016; 126(6):2334-40 [1]

Chapter 3: Panda A, Betigeri A, Subramanian K, et al. JCO Precision Oncology 2017 [2]

Chapter 4: Panda A, Mehnert JM, Hirshfield KM, et al. J Natl Cancer Inst. 2018; 110(3) [3]

Work of others

Figure 1.1, Figure 2.1, part of Figure 3.8, and Figure 4.3 are not my work. I compiled Figure 1.1 from literature with permissions, because it helped me explain the basic concepts relevant to my work. The mutation signatures depicted in Figure 3.8 is from COSMIC (`http://cancer.sanger.ac.uk/cosmic/signatures`), I included them because it helped me explain the protocol described in that figure. Figure 2.1 and Figure 4.3 are the works of my collaborators at the Rutgers Cancer Institute of New Jersey (R-CINJ) and the Foundation Medicine (Cambridge, MA), that I reproduced from the publications referenced above (of which I am an author or a coauthor) because the content of these figures directly support the findings of my analysis.

Source of data

List of genomic regions assessed by StrandAdvantage and FoundationOne assays were provided by Strand Life Sciences (Bangalore, India) and Foundation Medicine respectively, and I used this

information in the analysis presented in panel A of Figure 3.5. I used response data provided by Dr. Janice Mehnert from R-CINJ and Dr. Douglas Johnson from Vanderbilt Ingram Cancer Center (VICC) in the analysis presented in panel B, C, D of Figure 3.5. Mutation data generated by Foundation Medicine, and subsequently provided by Prof. Hossein Khiabanian (R-CINJ) and Dr. Douglas Johnson (VICC), was used in the analysis presented in panel E of Figure 3.2, and panel B, C, D of Figure 3.5. Panel D of Figure 2.2, the bottom row of Figure 3.3, and panel D of Figure 4.1 uses pathology based lymphocyte infiltration scores provided by Dr. Hua Zhong (then at R-CINJ), Dr. Anil Betigeri (then at Strand Life Sciences), and Dr. Greg Riedlinger (R-CINJ) respectively.

Gratitudes

I thank my advisor Prof. Gyan Bhanot, for accepting me in his computational biology group when I had experience in neither computation nor biology, for having faith in me and not giving up on me when I was unsure myself whether my decision to work on cancer was a good one or a fatal mistake, for being patient with me as he guided me from scratch, for always being available to listen to my ideas and to provide useful advises, for having helpful collaborators who provided data for validation, and for helping me with 'common sense' (which I often find counterintuitive). I thank my co-advisor Dr. Shridar Ganesan, for introducing me to immune checkpoint therapy, for funding my graduate assistantship/fellowship which allowed me to focus on my research instead of teaching, for finding time to listen to my ideas in spite of his busy schedule, for taking my ideas and results seriously in spite of my limited knowledge of biology/medicine, and for having valuable discussions with me that taught me interesting biology and helped me develop new ideas. I thank Prof. Manisha Bajpai for being always nice and supportive, and Jennifer Hostettler for her amazing editorial assistance in getting the above papers published. I thank my program directors, my thesis committee, Shirley Hinds and Jacqueline Harris, for the nice time I had at Rutgers. I thank the entire crew of Mithaas (Piscataway) and Shivaji (Edison) for preparing and delivering the food that kept me alive.

Dedication

To my parents, who put my education above all else;

and my paternal grandparents, who died of cancer.

Table of Contents

List of Figures

Chapter 1

Biomarkers of response to immune checkpoint therapy

1.1 Motivation

According to a 2016 report released by the Centers for Disease Control and Prevention (CDC) and acknowledged by the American Heart Association (AHA), cancer will soon surpass heart disease as the leading cause of death in the United States. The American Cancer Society (ACS) estimates that in 2017, three people were diagnosed with cancer and one person died of cancer every minute in the United States. Worldwide, 14 million were diagnosed with cancer and 8 million died of cancer in 2012, and these numbers are projected to rise to 21 million and 13 million respectively by 2030. The vast majority (over 90%) of cancer deaths are from metastasis [4, 5], which had led to the view that once a cancer metastasizes, the prognosis is poor and treatment options are limited. This is the reason for the intense focus on early detection, because of the belief that only cancers identified before metastasis are curable. Recently however, there has been a breakthrough in the treatment of metastatic cancers, because of a novel type of treatment called immune checkpoint therapy which has demonstrated unprecedented response in many patients with metastatic cancers from many different tissues of origin. Immune checkpoint therapy not only shrinks tumors but also has led to long term response in several cancers [6].

In spite of these substantial advances, durable response to immune checkpoint therapy is seen in only a subset of patients [6], and the mechanisms of response are not well understood. This was the motivation for my doctoral research. Some of the key questions I try to answer in this thesis are: Is it possible to analyze large public datasets, such as The Cancer Genome Atlas (TCGA), to identify potential biomarkers of response to immune checkpoint therapy? Can these potential biomarkers be

validated using data from patients treated with immune checkpoint therapy? Can these results be translated into clinical use? Do the clinically validated biomarkers reveal underlying mechanisms of response? Can this understanding be translated into improvements in standard of care?

1.2 Basic concepts

1.2.1 Immune recognition of tumor

The primary task of our immune system is to distinguish cells that are self from those that are non-self (e.g. pathogens, cells infected by pathogens etc.), and to destroy the non-self cells. There are many ways in which a tumor cell can be recognized as non-self by the immune system. For example, as we will see in Chapter 4, virally mediated tumors can be recognized by the immune system as non-self, probably because they express exogenous viral protein fragments on their cell surface which can make them look like virus infected cells to the immune system. Similarly, as we will see in Chapter 5, tumors expressing endogenous retroviruses (viral elements that are present in our own genome but are suppressed in normal somatic tissues) can also be sometimes recognized by the immune system as non-self. In addition to these, there are several other mechanisms that can lead to immune recognition of the tumor as non-self.

Some mutations (non-synonymous) in DNA can change the mRNA sequences when the DNA is transcribed, and subsequently the amino acid sequences of the protein when the mRNA is translated (panel A in Figure 1.1). Thus, tumor cells that have non-synonymous mutations in their DNA (i.e. mutations that change the amino acid sequence) will make proteins that are different from those in normal cells that lack these mutations. Whether the immune system can detect these novel non-self protein fragments (neo-antigens) or not, i.e. whether a neo-antigen is immunogenic or not, depends on a variety of complex biological factors. However, it seems logical to expect that the higher the number of non-synonymous mutations in a tumor, the higher the number of neo-antigens in the tumor, and the higher the probability that the tumor harbors immunogenic neo-antigens that the immune system can identify. So, hyper-mutation in a tumor, due to various causes (etiologies), may be a mechanism of immune recognition, as we will see in Chapters 2 and 3.

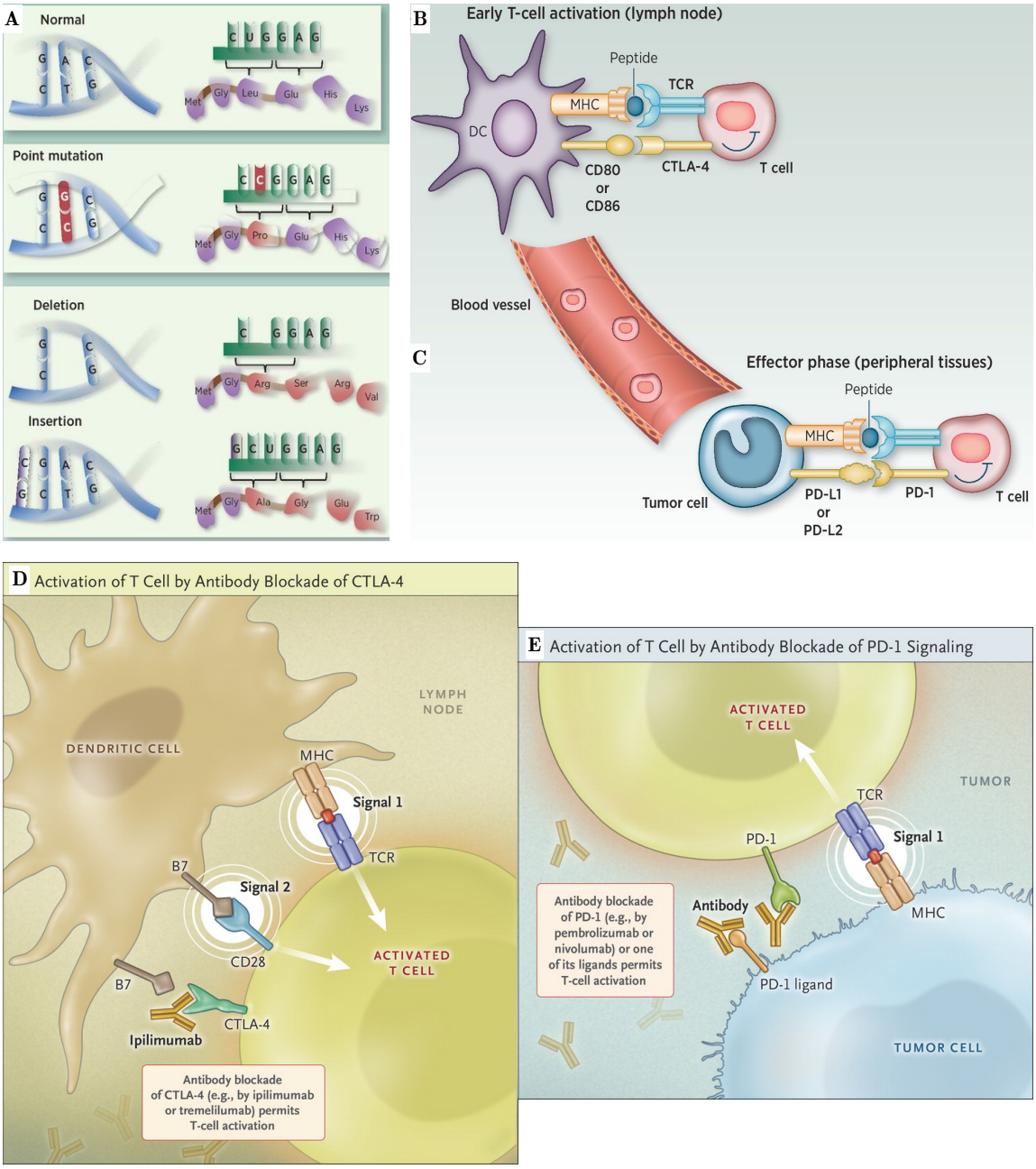


Figure 1.1: A mechanism of response to immune checkpoint therapy.

(A) Non-synonymous mutation in the DNA changes the sequence of the resulting mRNA, making the resulting protein different from its normal counterpart (figure reprinted from [7] with permission from American Association for Cancer Research). Thus a sufficiently high non-synonymous mutation burden in a tumor can make the tumor sufficiently non-self, leading to immune recognition. (B) Binding of CTLA-4 with CD80/CD86 (after immune recognition) or (C) binding of PD-1 with PD-L1/PD-L2 (after immune infiltration) can block the anti-tumor immune response led by CD8+ T cells, preventing CD8+ T cells from killing tumor cells (figures reprinted from [8] with permission from American Association for Cancer Research). (D) CTLA-4 blockade drugs bind with CTLA-4 preventing it from binding with CD80/CD86, whereas (E) PD-1 blockade drugs bind with either PD-1 or PD-L1 preventing them from binding with each other (figures reproduced with permission from [9], ©Massachusetts Medical Society). In this way, immune checkpoint therapy (e.g. CTLA-4 / PD-1 blockade) frees the blocked immune response, allowing CD8+ T cells to kill the tumor cells.

1.2.2 Immune checkpoint pathways

The immune system is composed of many different cell types, which includes cells like CD8+ T cell that can kill cells that it perceives as non-self. To prevent auto-immunity, i.e. to protect our normal cells from being accidentally killed by CD8+ T cells, there exist specific pathways, called immune checkpoint pathways, that are used as brakes on CD8+ T cells. Thus, CD8+ T cell response is regulated by multiple stimulatory and inhibitory signals [10]. Following immune recognition of tumor as non-self, tumors sometimes hijack the immune checkpoint pathways to block the anti-tumor immune response, thereby preventing CD8+ T cells from killing tumor cells. For example, in the early stage of immune activation, when CD8+ T cells are being trained to identify the specific neo-antigen that caused immune recognition, the immune checkpoint receptor CTLA-4 (expressed on the surface of CD8+ T cells) can bind to its ligands CD80 or CD86 (expressed on the surface of dendritic cells), which would send an inhibitory signal to CD8+ T cells, thereby preventing immune response (panel B of Figure 1.1). Alternatively, in later stage of immune activation, i.e. after CD8+ T cells have travelled to and infiltrated the tumor cells, the immune checkpoint receptor PD-1 (expressed on the surface of CD8+ T cells) can bind to its ligands PD-L1 or PD-L2 (expressed on the surface of tumor cells), which will send an inhibitory signal to CD8+ T cells preventing them from killing tumor cells (panel C of Figure 1.1).

1.2.3 Immune checkpoint therapy

As explained above, several mechanisms such as hyper-mutation in tumor due to various etiologies, and expression of exogenous or endogenous viral RNA in tumor, can cause immune recognition of a tumor, leading to anti-tumor immune response spearheaded by CD8+ T cells. Such tumors often co-opt the immune checkpoint pathways, whose normal function is to prevent auto-immunity, to block the immune response and to prevent CD8+ T cells from killing tumor cells. However, the same block also makes these tumors vulnerable to immune checkpoint disruption. In other words, if we can find a way to prevent the tumor from blocking the immune response, CD8+ T cells will eliminate the tumor cells. This is the idea behind immune checkpoint therapy. The mechanism of

action of immune checkpoint therapy is to disrupt this block on the immune response. For example, in CTLA-4 blockade (panel D of Figure 1.1), the drug binds to the receptor CTLA-4 expressed on the surface of CD8+ T cells, preventing CTLA-4 from binding with its ligands CD80 or CD86. Similarly, in PD-1 blockade (panel E of Figure 1.1), the drug binds to either the receptor PD-1 expressed on the surface of CD8+ T cells, or the ligand PD-L1 expressed on the surface of tumor cells, preventing the binding between PD-1 and PD-L1. By preventing the binding between immune checkpoint receptors and their ligands, immune checkpoint therapy undoes the tumor induced block on the immune response, allowing CD8+ T cells to kill the tumor cells.

1.3 Context of my research

Tumors from responders to PD-1 blockade in skin melanoma [11] were known to have a robust CD8+ T cell infiltration in the tumor as well as over-expression of immune checkpoint protein PD-L1 (a ligand of the receptor PD-1). This suggested the hypothesis that some mechanism in the tumor cells leads to immune recognition of these tumors causing CD8+ T cell infiltration in the tumor, which is then blocked by the tumor by over-expressing PD-L1, which in turn makes the tumors vulnerable to PD-1 blockade. It was subsequently shown that tumors from responders to CTLA-4 blockade in skin melanoma [12, 13] have significantly higher non-synonymous mutation burden compared to tumors from non-responders, implicating hyper-mutation in tumor as the mechanism of immune activation in skin melanoma. This finding was subsequently also corroborated by clinical data of PD-1 blockade in skin melanoma [14]. Apart from skin melanoma, mutation burden in the tumor was also found to be associated with response to PD-1 blockade in non-small cell lung cancer [15], colorectal cancer [16], and bladder urothelial cancer [17]. However, the mutation burden threshold that is necessary for immune activation and checkpoint pathway up-regulation in these cancer types was unknown. Knowledge of this threshold is essential to identify patients likely to respond to immune checkpoint therapy. It was also not known whether a similar mutation burden threshold exists in other solid cancer types which can identify patients with a blocked immune response, who are likely to benefit from immune checkpoint therapy. These questions will be addressed in Chapters 2 and 3.

Apart from such hyper-mutation related response, certain virally mediated tumors, such as Merkel cell polyoma virus (MCPyV) in Merkel cell cancer [18, 19] and Epstein-Barr virus (EBV) in NK/T cell lymphoma [20], were known to respond to PD-1 blockade, in spite of having a low mutation burden. However, it was unknown whether other virally mediated tumors are also likely to respond to immune checkpoint therapy. In addition, it was also unknown whether tumors expressing specific endogenous retrovirus (hERV) may also respond to immune checkpoint therapy in some cancer types. These questions will be addressed in Chapters 4 and 5.

1.4 Summary of the results

Using TCGA (a dataset of >10,000 tumors from >30 cancer types) as discovery dataset, recently published datasets of patients treated with immune checkpoint therapy for retrospective validation, and data of patients treated with immune checkpoint therapy at the Rutgers Cancer Institute of New Jersey (R-CINJ) and the Vanderbilt Ingram Cancer Center (VICC) for prospective validation, we identified two potential biomarkers of response to immune checkpoint therapy, as briefly described below, and then elaborated in subsequent chapters.

1.4.1 Hyper-mutation in tumor due to various etiologies

A. Hyper-mutation due to POLE proof-reading defect (Chapter 2)

(Previously published work: Mehnert JM, Panda A, Zhong H, et al. Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. J Clin Invest. 2016; 126(6):2334-40 [1])

Analysis of the TCGA endometrial cancer dataset showed that a POLE mutation in tumor was associated with an ultra-mutator phenotype, elevated CD8+ T cell infiltration, over-expression of immune response genes, and up-regulation of immune checkpoint pathways. This suggested that hyper-mutation in tumor due to POLE proof-reading defect is a potential biomarker of response to immune checkpoint therapy. This prediction was supported by the case of a patient with POLE-mutant metastatic endometrial cancer at R-CINJ who had an exceptional response to PD-1 blockade.

B. Hyper-mutation due to other mutational etiologies (Chapter 3)

(Previously published work: Panda A, Betigeri A, Subramanian K, et al. Identifying a clinically applicable mutation burden threshold as a biomarker of response to Immune Checkpoint Therapy in solid tumors. JCO Precision Oncology 2017 [2])

9,472 tumors in 33 solid cancer types from TCGA were analyzed to determine whether an immune Checkpoint Activating Mutation burden threshold (iCAM) associated with evidence of immune activation and checkpoint pathway upregulation exists in any of these cancers. This analysis showed that, in addition to the 4 known cancers mentioned above, such a threshold exists in 4 more cancers: endometrial, gastric, cervical, and ER+ HER2− breast cancer. Blinded pathological evaluation confirmed the presence of a robust lymphocytic infiltration in iCAM+ tumors (tumors with mutation burden above the threshold) in the TCGA dataset. In publicly available datasets of CTLA-4 blockade in skin [12, 13], and PD-1 blockade in lung [15] and colon [16] cancer, the iCAM threshold was able to distinguish between responders and non-responders with high accuracy. ROC analysis in TCGA dataset showed that iCAM+ tumors can be identified with high accuracy using routine clinical sequencing assays (e.g. FoundationOne). Finally, analysis of 113 skin melanoma sequenced using the FoundationOne clinical assay showed that iCAM+ patients had significantly better response to PD-1 blockade. In the TCGA dataset, iCAM+ tumors had mutational etiologies [21, 22] associated with UV exposure or smoking in skin and lung cancer respectively, mismatch repair defect or POLE proof-reading defect in colon / endometrial / gastric cancer, and aberrant APOBEC activity (an enzyme whose function is to mutate and disable viruses that infect cells, but can accidentally mutate host genome as well) in bladder / cervical / ER+ HER2− breast cancer. In contrast, iCAM− tumors had a stronger contribution from deamination of 5-methylcytosine, indicative of aging [23].

1.4.2 Expression of exogenous or endogenous viral RNA in tumor

A. Expression of exogenous virus in tumor (Chapter 4)

(Previously published work: Panda A, Mehnert JM, Hirshfield KM, et al. Immune activation and benefit from avelumab in EBV-positive gastric cancer. J Natl Cancer Inst. 2018; 110(3) [3])

Immune checkpoint therapy is known be effective in a subset of gastric cancer patients with low tumor mutational burden. To understand the mechanism underlying this, we analyzed gene expression and mutation data from the TCGA gastric cancer cohort and found that tumors expressing EBV had elevated CD8+ T cell infiltration, over-expression of immune response genes, and up-regulation of immune checkpoint pathways. This suggested that EBV+ gastric cancer may respond to immune checkpoint therapy. This prediction was supported by the case of a patient with EBV+ metastatic gastric cancer at R-CINJ who had durable clinical benefit from PD-1 blockade.

B. Expression of endogenous retroviruses in tumor (Chapter 5)

(Unpublished work: manuscript is currently under preparation)

Analysis of TCGA clear-cell renal cancer (ccRCC) dataset using published hERV expression data [24] identified 20 hERVs whose expressions were correlated with immune activation and checkpoint pathway up-regulation. Analysis of 20 more cancer types from TCGA showed similar evidence of hERV-associated immune checkpoint activation in ER+ HER2− breast, colon, and head-neck squamous cell cancers. Only two hERVs, ERVK-2 and ERV3-2, were associated with immune checkpoint activation in all 4 cancer types. ERV3-2 expression was correlated with immune checkpoint activation in 7 more cancer types (i.e. >10 solid cancer types in total). Clinical validation of ERVK-2 and ERV3-2 expressions as biomarkers of response to PD-1 blockade in ccRCC is currently in progress.

1.5 Basic methods

1.5.1 Processing of gene expression data

RNAseqV2 scaled estimates were obtained from Broad GDAC (`http://gdac.broadinstitute.org`) and TCGA data portal (`https://tcga-data.nci.nih.gov`). Scaled estimates were then median-adjusted to eliminate systematic error, so that median scaled estimate is 1 in every sample. This median-adjusted RNA-seq data was used as input for ESTIMATE [25] and CIBERSORT [26] algorithms, to quantify the level of immune infiltration in tumor, and the composition of tumor infiltrating leukocytes (respectively). Only un-ambiguous ($P < 0.05$) CIBERSORT outputs were

used in the analyses. For ease of visualization, the median-adjusted RNA-seq data (x) was then log-transformed to y = log_2(1+ 1023 x) so that x = 0 maps to y = 0 and the median expression is y = 10 in each sample. The factor 1023 (i.e. $2^{10} - 1$) was chosen because the distribution of positive x values in $log2$ scale was mostly in the range $(-10, 10)$.

1.5.2 Processing of somatic mutation data

For Chapter 2, all Mutation Annotation Format (MAF) files for endometrial cancer as of 30th September 2015 were downloaded (`https://wiki.nci.nih.gov/display/TCGA/TCGA+MAF+Files`), re-annotated using Oncotator [27], merged and de-duplicated. For subsequent Chapters, all MAF files (both public and protected) current as of 31st January 2016 were downloaded, mapped from hg18 to hg19 using liftOver (`https://genome.ucsc.edu/cgi-bin/hgLiftOver`) if necessary, re-annotated using Oncotator [27], merged and de-duplicated.

1.5.3 Collection of clinical data

Clinical data, including microsatellite status and human papillomavirus (HPV) status, were obtained from TCGA data portal. Luminal-A/B status of breast cancer samples, and EBV status of gastric cancer samples were obtained from corresponding TCGA publications [28, 29]. ERBB2 focal copy number data and ESR1 mRNA expression data from Broad GDAC were used to classify breast cancer into clinical subtypes (ER+ HER2−, ER− HER2−, HER2+) that were analyzed separately.

1.5.4 Software and statistical tests used

MATLAB and R were used for all analysis. Apart from CIBERSORT [26] P-values, Fisher's test and log-rank test were used to compare response rates and survival data respectively, and Wilcoxon rank-sum test was used for all other comparisons. Spearman rho was used for all test of correlation. In all cases, statistical significance was assessed at $P < 0.05$ using two-sided tests.

Chapter 2

Hyper-mutation due to proofreading defect as a biomarker

Published as: Mehnert JM, Panda A, Zhong H, et al. J Clin Invest. 2016; 126(6):2334-40 [1]

2.1 Introduction

The human body consists of tens of trillions of cells, and each cell has billions of base pairs of DNA. When cells divide, this DNA is copied (replication), and mistakes (mutations) can be made in this copying process. Various DNA repair machineries, such as proofreading (corrections during copying) and mismatch repair (corrections after copying), are in place to ensure that the copying is as accurate as possible. However, if these DNA repair machineries are compromised, the mutation rate will skyrocket, as mutations will accumulate each time a cell divides.

Proofreading is done in general by a group of enzymes called DNA polymerase, but 2 specific polymerases (polymerase epsilon and polymerase delta) are responsible for proofreading of DNA in eukaryotes, where the genes POLE (DNA polymerase epsilon, catalytic subunit) and POLD1 (DNA polymerase delta 1, catalytic subunit) play key roles by virtue of their exonuclease domain. Therefore, non-synonymous mutations in POLE or POLD1, especially those in the exonuclease domain, can lead to proofreading defect in a cell, causing extremely high mutation burden.

2.2 An exceptional response in endometrial cancer

A patient with metastatic endometrial cancer recently had an exceptional response to PD-1 blockade (panel A and B of Figure 2.1) in an IRB approved phase I clinical trial (pembrolizumab at 10 mg/kg every 2 weeks) at R-CINJ. The patient experienced rapid clinical improvement, with an objective

(partial) response after only 8 weeks of treatment, and the response was still ongoing at 14 months (her last follow-up). She tolerated the therapy well (only mild rash and liver function test elevation), and a moderate fever early in the course of treatment that resolved spontaneously.

2.2.1 Robust lymphocyte infiltration and over-expression of PD-L1

The patient was originally diagnosed with stage IB grade III endometrioid type endometrial adenocarcinoma at the age of 53, and histologic assessment of the primary tumor showed robust immune infiltration in the tumor (panel D of Figure 2.1). She initially deferred radiation therapy, but was treated with chemotherapy and radiotherapy after a rapid recurrence of cancer. Metastasis was detected 2 years after that, and histologic assessment of the metastatic tumor showed strong PD-L1 expression (panel E of Figure 2.1). She was then enrolled in this phase I trial of PD-1 blockade.

2.2.2 Ultra-mutator phenotype associated with proofreading defect

With her informed consent, her pre-treatment tumor samples (both primary and lymph node metastasis) were sent for clinical sequencing of all exons of 315 cancer-related genes (FoundationOne assay). The sequencing showed that both tumors had an ultra-mutator phenotype (panel C of Figure 2.1): 32 likely pathogenic variants and 116 variant of unknown significance (VUS) in the primary tumor, and 33 likely pathogenic variants and 159 VUS in the metastatic tumor, with 28 likely pathogenic variants and 83 VUS common to both, which are unusually high as endometrial tumors usually have only 7-8 likely pathogenic variants and 10-15 VUS. None of the likely pathogenic variants, and only 3 of the VUS were detected in sequencing of germline DNA, confirming most of the above variants to be somatic. Both primary and metastatic tumors harbored a mutation (V411L) in the exonuclease domain of POLE that affects DNA proofreading function of this gene, and a nonsense mutation (R114*) in POLE suggesting inactivation of the other copy. Query of POLE mutation in cBioPortal [30, 31] showed that V411L is a mutation hotspot in POLE (panel C of Figure 2.1). 23 out of 252 (9.1%) advanced endometrioid type endometrial tumors sequenced using FoundationOne assay had mutation in POLE, and POLE-mutant tumors had 21.2 $\pm$ 4.1 likely pathogenic variants and 82.2

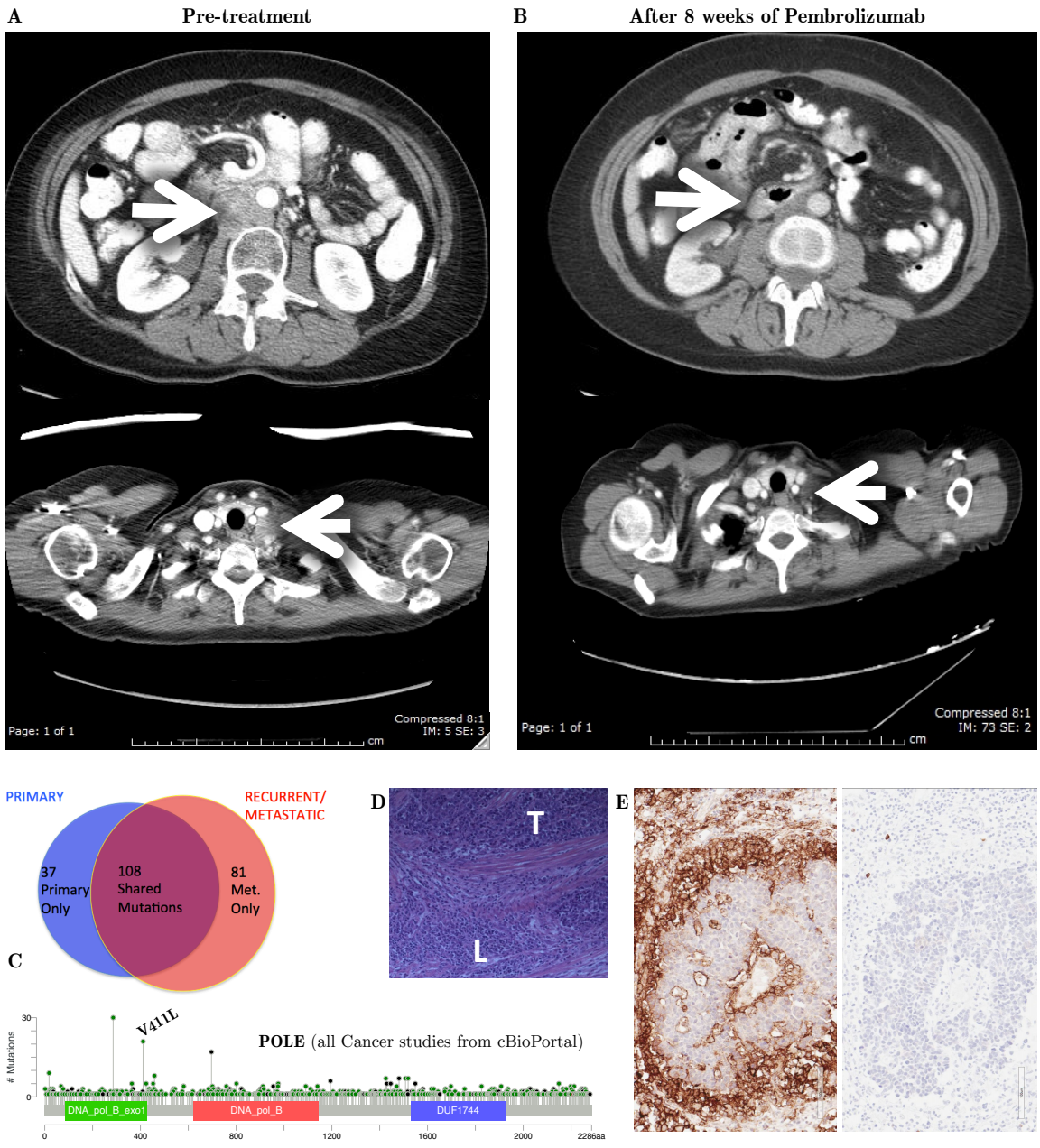


Figure 2.1: An exceptional response to PD-1 blockade in metastatic endometrial cancer

A patient with metastatic endometrial cancer (A) had an exceptional response to PD-1 blockade (B) at the Rutgers Cancer Institute of New Jersey. Sequencing of 315 cancer-related genes in primary and metastatic tumor of the patient showed >100 non-synonymous somatic mutations (C), including a hotspot mutation in the exonuclease domain of POLE, a gene that plays a crucial role in DNA proofreading. Histologic assessment showed that the tumor had robust lymphocyte infiltration (D; T: tumor, L: lymphocyte), and was strongly positive for PD-L1 (E; left: IHC staining, right: negative control). This figure was reprinted with permission from our previously published work [1].

± 25.0 VUS compared to 7.5 ± 0.5 likely pathogenic variants and 12.8 ± 2.6 VUS in POLE-wild-type tumors, suggesting an association between POLE mutation and ultra-mutator phenotype in advanced endometrial cancer. Similar associations has been reported in primary endometrial cancer [32], and yeast models have shown that mutations in exonuclease domain of POLE can indeed cause a mutator phenotype [33].

2.3 Results from TCGA endometrial cancer cohort

2.3.1 Operating hypothesis

The genomic and histologic observations in this case of exceptional response suggested the following hypothesis regarding the mechanism of response to immune checkpoint therapy in endometrial cancer: since POLE plays an important role in proofreading of DNA, tumors with POLE mutation have a proofreading defect that causes an ultra-mutator phenotype, as observed in the exceptional responder. The extremely high non-synonymous mutation burden in these tumors causes immune recognition of these tumors as non-self, leading to a robust immune infiltration in these tumors, as observed in the exceptional responder. To survive this anti-tumor immune response led by CD8+ T cells, these tumors are forced to use immune checkpoint pathways (e.g. PD-1 pathway) to block the immune response, consistent with the PD-L1 overexpression observed in the exceptional responder. Since these tumors depend on immune checkpoint pathways for survival, they are vulnerable to immune checkpoint disruption (e.g. PD-1 blockade), as observed in the exceptional responder.

2.3.2 Study design

Motivated by the above hypothesis, we analyzed the TCGA endometrial cancer dataset to find supporting evidence. Endometrial cancer has 2 histological subtypes, serous and endometrioid, where the former has an abundance of copy number alterations that the latter lacks. Due to this difference in patterns of alteration, and also because the exceptional responder had endometrioid type endometrial cancer, we excluded serous samples from the analysis and focused on the endometrioid samples only. A subset of endometrial cancer was known to have microsatellite instability [32], and

tumors with microsatellite instability was known to respond to PD-1 blockade in colorectal cancer [16], which suggested the possibility that tumors with microsatellite instability may respond to PD-1 blockade in endometrial cancer as well. Therefore, we classified the endometrioid samples into 3 groups: (i) POLE (samples with non-synonymous somatic mutation in POLE; N = 27), (ii) MSI (samples with microsatellite instability but no non-synonymous somatic mutation in POLE; N = 64), (iii) MSS (microsatellite stable samples with no non-synonymous somatic mutation in POLE; N = 104), and compared these groups pairwise for various immunological markers of interest.

2.3.3 Mutation burden

MSI tumors are known to have mismatch repair defect, and consistently MSI tumors had significantly higher non-synonymous mutation burden compared to MSS tumors in TCGA endometrial cancer cohort. Interestingly, POLE mutant endometrial tumors had a significantly higher non-synonymous mutation burden compared to even MSI tumors (panel A of Figure 2.2). This was consistent with the hypothesis that POLE mutation is associated with an ultra-mutator phenotype.

2.3.4 Immune activation

POLE mutant tumors also had significantly higher mRNA expression of CD8A, a marker of CD8+ T cells, not just compared to MSS tumors, but even compared to MSI tumors, suggesting an abundance of CD8+ T cell infiltration in POLE mutant tumors (panel B of Figure 2.2).

For a more comprehensive evaluation, we compared the expression of around 700 immune-related genes among the 3 groups. Of these 16 genes were differentially expressed between POLE mutant tumors and MSI tumors at a 25% false discover rate (Benjamini Hochberg), and 111 genes were differentially expressed between POLE mutant tumors and MSS tumors at a 5% false discover rate (Benjamini Hochberg). The heatmap (panel C) in Figure 2.2 shows the relative expression of these differentially expressed immune-related genes (rows) in the tumors (column), where the top quartile is colored gold and the bottom quartile is colored cyan. This demonstrated that POLE-mutant tumors (red) had high expression of a large set of immune-related genes compared to MSS tumors

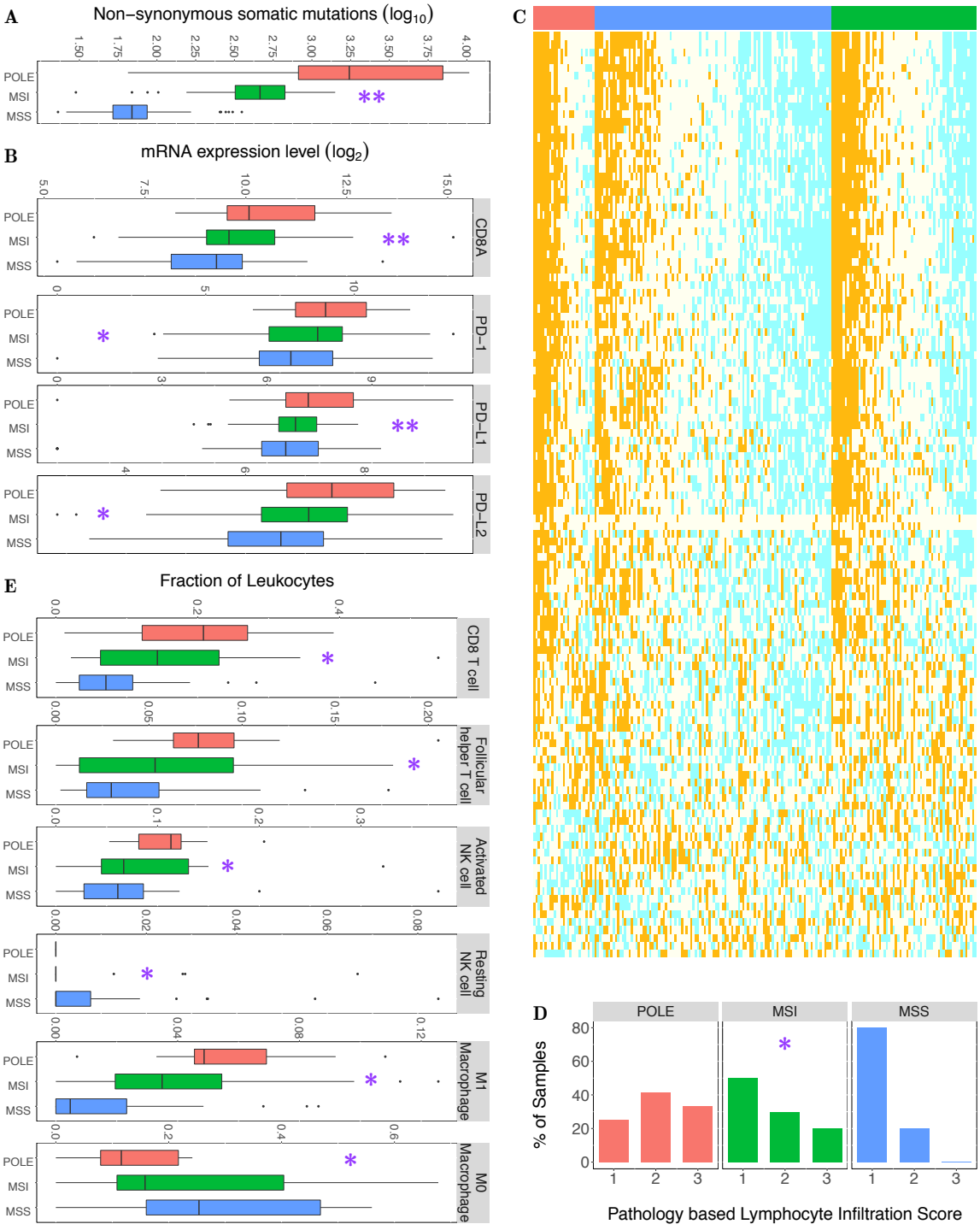


Figure 2.2: Comparison of POLE-mutant, MSI, MSS tumors in TCGA endometrial cancer cohort. (A) Number of non-synonymous mutations. (B) mRNA expression of cytotoxic T cell marker CD8A, and PD-1 checkpoint pathway genes. (C) Heatmap of relative expression of differentially expressed immune-related genes (gold: high, cyan: low). (D) Distribution of pathology based lymphocyte infiltration scores (in 12 POLE, 10 MSI, 10 MSS tumors; 1: low, 3: high). (E) Fractional composition of tumor-infiltrating leukocytes. ** both POLE vs MSI and POLE vs MSS differences are statistically significant, * POLE vs MSS difference is statistically significant, at $P < 0.05$ in two-sided Wilcoxon ranksum test. This figure was reprinted with permission from our previously published work [1].

(blue), while MSI tumors (green) had an intermediate phenotype.

As a consistency check, high resolution digital images of hematoxylin and eosin (H&E) stained formalin-fixed paraffin-embedded (FFPE) sections of 12 POLE mutant, 10 MSI, 10 MSS tumors from TCGA endometrial cancer cohort were evaluated by a pathologist, blinded to genomic data. The pathologist scored each image on a scale of 1 (minimal) to 3 (robust) for the presence of tumor infiltrating lymphocytes. The distribution of this pathology based lymphocyte infiltration score in the 3 groups (panel D of Figure 2.2) showed that POLE-mutant tumors had a significantly higher lymphocytic infiltration compared to MSS tumors, while MSI tumors had an intermediate phenotype, which was consistent with our findings from RNA-seq data. This was also consistent with the findings of a recent pathological study that reported lymphocyte infiltration in POLE mutant endometrial cancer [34].

These results confirmed that in agreement with our hypothesis, POLE mutation is indeed associated with immune activation in endometrial cancer.

2.3.5 Checkpoint pathway up-regulation

Compared to MSS tumors, POLE mutant tumors also had a significantly higher mRNA expression of the immune checkpoint receptor PD-1 and its two ligands PD-L1 and PD-L2 (panel B of Figure 2.2), indicating up-regulation of the PD-1 pathway in POLE mutant tumors. Moreover, POLE mutant tumors had a significantly higher mRNA expression of the ligand PD-L1 compared to even MSI tumors (panel B of Figure 2.2). Since higher PD-L1 expression in solid tumors is associated with better response to PD-1 blockade [35], this particular result suggests that POLE mutant endometrial tumors may respond to PD-1 blockade even better than MSI endometrial tumors. These results further confirmed our expectation of POLE mutations being associated with checkpoint pathway up-regulation in endometrial cancer.

2.3.6 Immune microenvironment

Additional analysis of the 3 groups showed that tumors in each group had different compositions of tumor infiltrating leukocytes. CD8+ T cells constituted a significantly higher fraction of tumor infiltrating leukocytes in POLE mutant tumors (N = 12) compared to MSS tumors (N = 28), while MSI tumors (N = 20) had an intermediate phenotype. This showed that a POLE mutation is associated with elevated CD8+ T cell infiltration in endometrial cancer, which is also consistent with the above hypothesis. This was also consistent with the finding of a recent study that reported robust CD8+ T cell response in an independent cohort of 47 POLE mutant endometrial cancer [36].

The relative abundance of follicular helper T cells was also significantly higher in POLE mutant tumors compared to MSS tumors, with MSI tumors having an intermediate phenotype. This was consistent with the finding of a recent study that POLE mutation induces robust tumor-specific CD4+ T cell response in endometrial cancer [37]. Apart from T cells, interesting patterns were also observed for natural killer (NK) cells and macrophages. The relative abundance of activated NK cells was significantly higher in POLE mutant tumors compared to MSS tumors, but the relative abundance of resting NK cells was significantly lower in POLE mutant tumors compared to MSS tumors, which suggest that POLE mutation in tumor is associated with activation of NK cells. Similarly, the relative abundance of M1 polarized macrophages was significantly higher in POLE mutant tumors compared to MSS tumors, but the relative abundance of M0 polarized macrophages was significantly lower in POLE mutant tumors compared to MSS tumors, which probably suggests that POLE mutation in tumor is associated with M0 to M1 transition of tumor infiltrating macrophages. Since NK-cells have important anti-tumor activity [38] and M1-macrophages can curtail tumor growth by suppressing angiogenesis and inducing apoptosis [39], these results suggest that POLE mutant tumors may have a more favorable immune microenvironment. Hence, in such tumors, if the brake on the immune system is removed via immune checkpoint therapy, it may more effectively eliminate the cancer cells.

2.3.7 Conclusion

To summarize, the results from TCGA endometrial cancer cohort showed that POLE mutations in tumor are associated with an ultra-mutator phenotype, immune activation, and checkpoint pathway up-regulation in endometrial cancer. A recent study reached the same conclusion in an independent cohort of 63 endometrial cancer patients [40]. These results support the above hypothesis, and suggests that POLE mutant tumors that constitute around 10% of endometrial cancer may be sensitive to immune checkpoint therapy, possibly even more sensitive than MSI tumors in endometrial cancer. Cases of exceptional response to PD-1 blockade in POLE mutant hyper-mutant tumors have recently also been reported in colorectal cancer [41], glioblastoma [42], and pediatric glioblastoma [43]. Thus, hyper-mutation in tumor due to POLE proofreading defect may be a biomarker of response to immune checkpoint therapy in endometrial cancer and other solid cancer types.

Chapter 3

Hyper-mutation due to some other etiologies as a biomarker

Published as: Panda A, Betigeri A, Subramanian K, et al. JCO Precision Oncology 2017 [2]

3.1 Introduction

Tumors from responders to CTLA-4 blockade in melanoma [12, 13] and PD-1 blockade in lung [15], colorectal [16], urothelial [17] cancers have significantly higher mutation burden than tumors from non-responders. This observation has led to the hypothesis that a sufficiently high non-synonymous mutation burden in a tumor may lead to immune recognition of the tumor as non-self, triggering an anti-tumor immune response spearheaded by CD8+ T cells. Some of these tumors block this response using immune checkpoint pathways, which makes these tumors vulnerable to immune checkpoint therapy. Although an association between mutation burden and response to immune checkpoint therapy was previously known in the above cancer types, it was not clear whether or not this relationship exists in other solid cancer types. Moreover, an immune checkpoint activating mutation burden threshold (iCAM) that can optimally identify patients likely to respond to immune checkpoint therapy in each cancer type was also unknown. Using RNA-seq based expression data and whole exome sequencing (WXS) based somatic mutation data for all solid cancer types in TCGA, we asked the following questions: (1) In which cancer types there is a mutation burden threshold above which tumors display a signature of immune activation and checkpoint pathway up-regulation? (2) Can this threshold distinguish responders from non-responders to immune checkpoint therapy? (3) Can this threshold be identified using routine clinical sequencing assays that interrogate only a few hundred genes? (4) Do tumors above and below this threshold have different mutational and immunological properties?

3.2 Methods

3.2.1 Determination of iCAM threshold in TCGA

If a non-synonymous mutation burden is associated with immune checkpoint activation in a cancer type, it should be possible to find a mutation burden threshold above which tumors have immune activation (led by CD8+ T cells) and checkpoint pathway upregulation. Figure 3.1 shows the method used to find the threshold. We used 3 sub-criteria – (a) mRNA level of CD8A (a marker of CD8+ T cells), (b) overall immune infiltration in tumor (quantified using ESTIMATE algorithm [25]), and (c) CD8+ T cell fraction in infiltrating leukocytes (quantified using CIBERSORT algorithm [26]) – to identify immune activation; and 6 sub-criteria – mRNA level of (d) checkpoint receptor PD-1 and its ligands (e) PD-L1 and (f) PD-L2, and (g) checkpoint receptor CTLA-4 and its ligands (h) CD80 and (i) CD86 – to identify checkpoint pathway upregulation.

In a cancer type with $N+1$ unique non-synonymous mutation burden values ($\{x_0, x_1, ..., x_N\}$ in ascending order), for each x_i in $\{x_1, ..., x_N\}$, we checked whether tumors with mutation burden $\geq x_i$ had significantly stronger immune activation and checkpoint pathway up-regulation compared to tumors with mutation burden $< x_i$ (as explained in Figure 3.1). In 8 solid cancer types, both criteria were simultaneously satisfied for one or more x-values, confirming an association between mutation burden and immune checkpoint activation in those cancer types. As shown in Figure 3.1 for 3 known cancer types, the x-values that satisfied a maximum number of sub-criteria were selected and ranked by the P-value of each satisfied sub-criterion, and the optimal x-value (i.e. the iCAM threshold) was identified by minimizing the sum of ranks over these sub-criteria.

3.2.2 Projection of iCAM threshold from TCGA to validation studies

To stratify tumors into iCAM+ and iCAM− in validation datasets, it was necessary to project the iCAM threshold from TCGA dataset to validation datasets. Distribution of the non-synonymous mutation burden (in log_{10} scale) in TCGA and a WXS based study (S) of the same cancer type were first aligned as follows: For unimodal distributions with peaks at x_1 and y_1 in TCGA and S

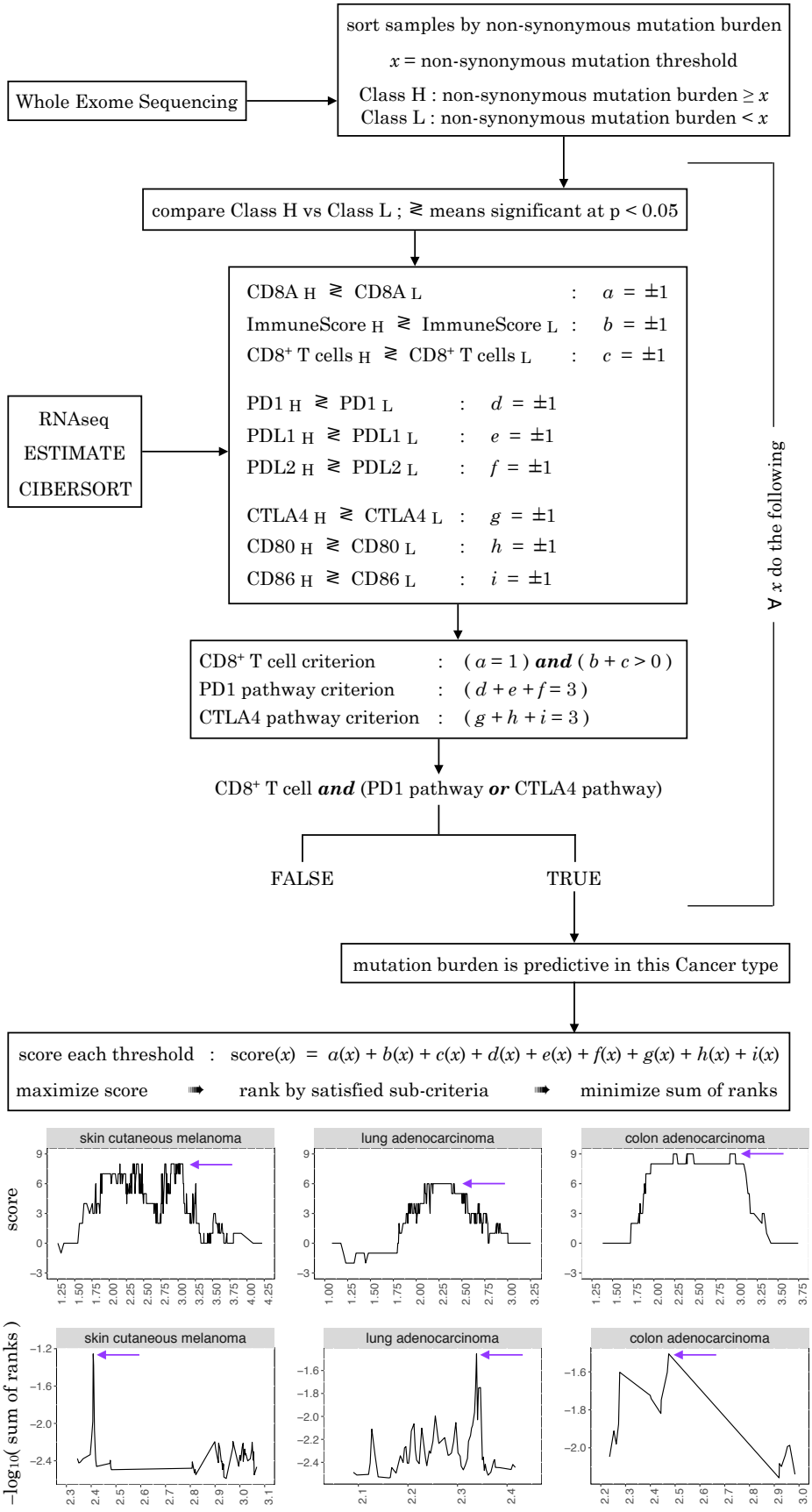


Figure 3.1: Protocol for identifying iCAM threshold in TCGA dataset.
This figure was reprinted with permission from our previously published work [2].

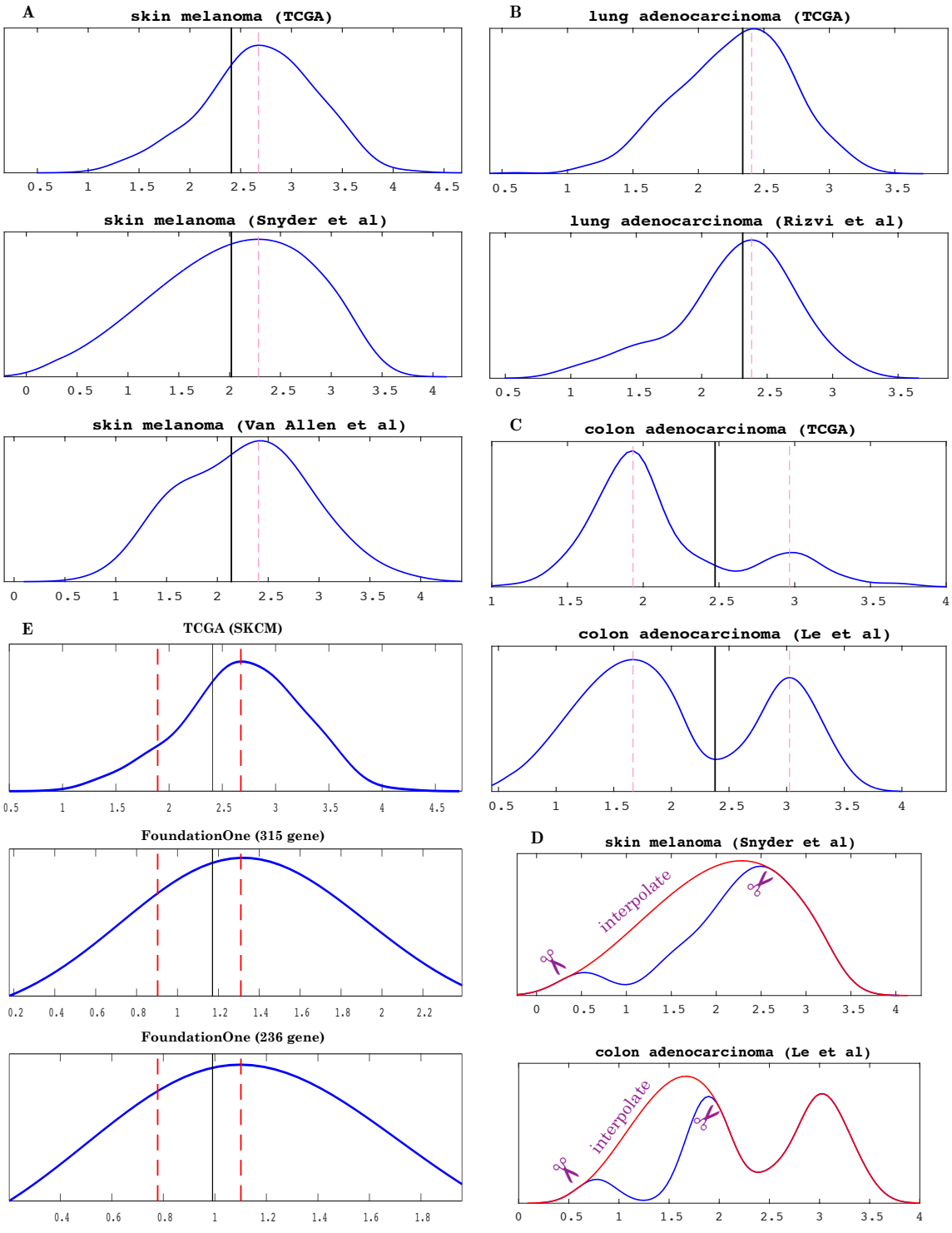


Figure 3.2: Protocol for projecting iCAM threshold from TCGA dataset to validation datasets. For threshold projection, mode(s) of WXS datasets are aligned with mode(s) of corresponding TCGA cohort through translation (A, B) or translation + scaling (C). Extra modes are smoothed out using interpolation (D) before alignment. For clinical sequencing assays, translation + scaling, using the mode and a reference point (e.g. 10 percentile), is needed to align even unimodal distributions (E). This figure was reprinted with permission from our previously published work [2].

respectively, the mutation axis of S was translated by $(x_1 - y_1)$, as shown in panel A and B of Figure 3.2. For bimodal distributions with peaks at (x_1, x_2) and (y_1, y_2) in TCGA and S respectively, the mutation axis of S was translated by $-b$ and then scaled by $1/a$, where the line $y = ax + b$ passes through (x_1, y_1) and (x_2, y_2), as shown in panel C of Figure 3.2. With the distributions aligned, the iCAM threshold (the black vertical line) can be mapped from TCGA data to S. This alignment requires that both datasets have the same number of peaks. If S had more peaks than TCGA (possibly due to small sample sizes), then S was smoothed using interpolation before alignment (panel D of Figure 3.2). When the mutation burden scale in S was very different from TCGA (e.g. $O(10^2)$ genes in clinical assays vs $O(10^4)$ genes in WXS), scaling was required in addition to translation even for unimodal distributions. In such cases, the distributions were aligned using the locations (x_1, y_1) of the peaks and the locations (x_2, y_2) of the 10th percentiles in the two datasets, using the procedure described above for bimodal distributions (panel E of Figure 3.2).

3.3 Results

3.3.1 Identification of the iCAM threshold

Using WXS based somatic mutation data and RNA-seq based gene expression data for 33 solid cancer types from TCGA, we looked (Figure 3.1) for a non-synonymous mutation burden threshold (iCAM), such that iCAM+ tumors (i.e. tumors with mutation burden above this threshold) had evidence of immune activation and checkpoint pathway upregulation. The top panel of Figure 3.3 shows the result of this analysis, where the sub-criteria satisfied are marked in gold in each cancer type. This analysis detected the presence of a robust iCAM threshold in 8 solid cancer types, which included the 4 known cancer types (skin melanoma [12, 13], lung adenocarcinoma [15], colon adenocarcinoma [16], and bladder urothelial cancer [17]) and 4 novel cancer types, namely endometrial cancer, stomach adenocarcinoma, cervical cancer, ER+ HER2− breast cancer. 3 other cancer types, namely serous ovarian cancer (OV), head-neck squamous cell cancer (HNSC), and prostate adenocarcinoma (PRAD), were also flagged in this analysis. However, these were not analyzed further because iCAM+ ovarian cancer had mutation burdens too low to be identified

accurately by clinical sequencing assays (that typically interrogate only a few hundred genes), and iCAM+ head-neck squamous and prostate cancers were too rare to be verified in clinical datasets (that typically have limited sample sizes).

The distribution of non-synonymous mutation burdens (in log_{10} scale) in the 8 cancer types, with the tumors stratified into iCAM+ and iCAM− classes, are shown in Figure 3.3 (middle). It should be noted that these thresholds are for the TCGA dataset, and specific to our use of the union of mutation calls from TCGA centers as the definition of mutation burden. For different definitions of mutation burden or different depths of sequencing, the thresholds would need to be recomputed. On a log_{10} scale, the iCAM threshold was 2.284 ± 0.205 in these cancer types. As a consistency check, high resolution images of H&E stained histological sections of 15 iCAM+ and 15 iCAM− tumors for each cancer type were pathologically evaluated in a blinded fashion for the presence of tumor infiltrating lymphocytes. Figure 3.3 (bottom) shows the distribution of pathology based lymphocyte infiltration scores in iCAM+ and iCAM− tumors in the 8 cancer types. As expected, iCAM+ tumors had significantly higher lymphocyte infiltration in all 8 cancer types.

3.3.2 iCAM status predicts response to PD-1 and CTLA-4 blockade in published data

To test whether iCAM status predicts response to immune checkpoint therapy in published clinical dataset, the iCAM threshold was projected from TCGA data to published datasets of patients treated with CTLA-4 blockade in skin melanoma [12, 13] (panel A of Figure 3.2), and PD-1 blockade in lung adenocarcinoma [15] (panel B of Figure 3.2) and colon adenocarcinoma [16] (panel C of Figure 3.2). In each dataset, tumors were stratified as iCAM+ or iCAM− based on this projected iCAM threshold. For consistency with the TCGA dataset, uveal melanoma samples were excluded from the Snyder [12] dataset, mucosal melanoma samples were excluded from the Van Allen [13] dataset, lung squamous samples were excluded from the Rizvi [15] dataset, and small bowel and ampulla of vater samples were excluded from the Le [16] dataset, before the threshold projection.

For CTLA-4 blockade in melanoma, in the Snyder dataset (panel A of Figure 3.4), around 60% (27/46) of iCAM+ patients had long-term benefit, compared to none (0/17) of iCAM− patients.

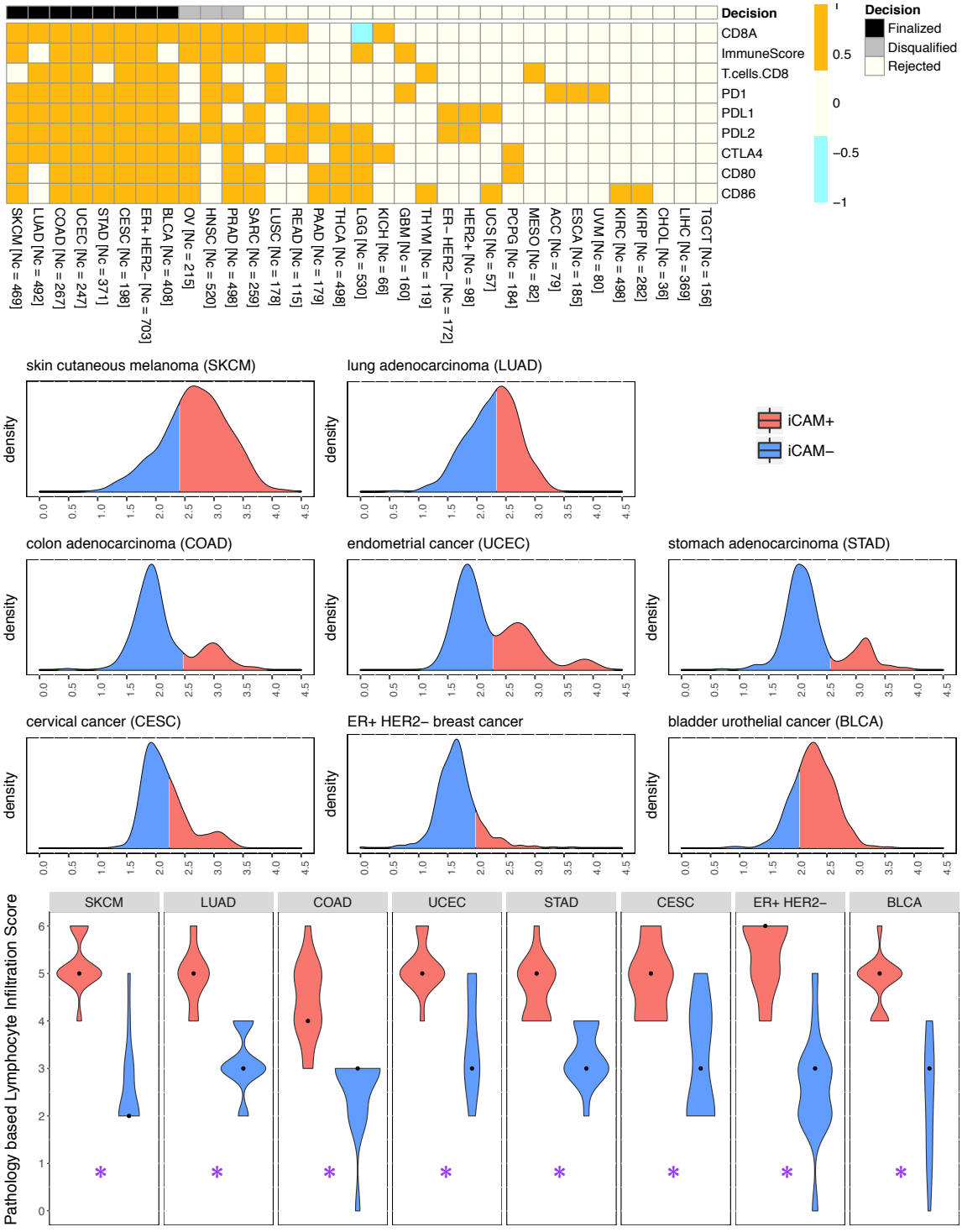


Figure 3.3: Mutation burden threshold associated with immune checkpoint activation in 8 cancers. Pan-cancer analysis of the TCGA dataset (top) identified iCAM threshold in 8 cancer types (middle). Gold: significantly higher in iCAM+ compared to iCAM−. Blinded assessment of 15 iCAM+ and 15 iCAM− tumors of each cancer type by a pathologist confirmed the presence of strong lymphocyte infiltration in iCAM+ tumors (bottom). * difference between iCAM+ and iCAM− is statistically significant at $P < 0.05$ in two-sided Wilcoxon ranksum test. This figure was reprinted with permission from our previously published work [2].

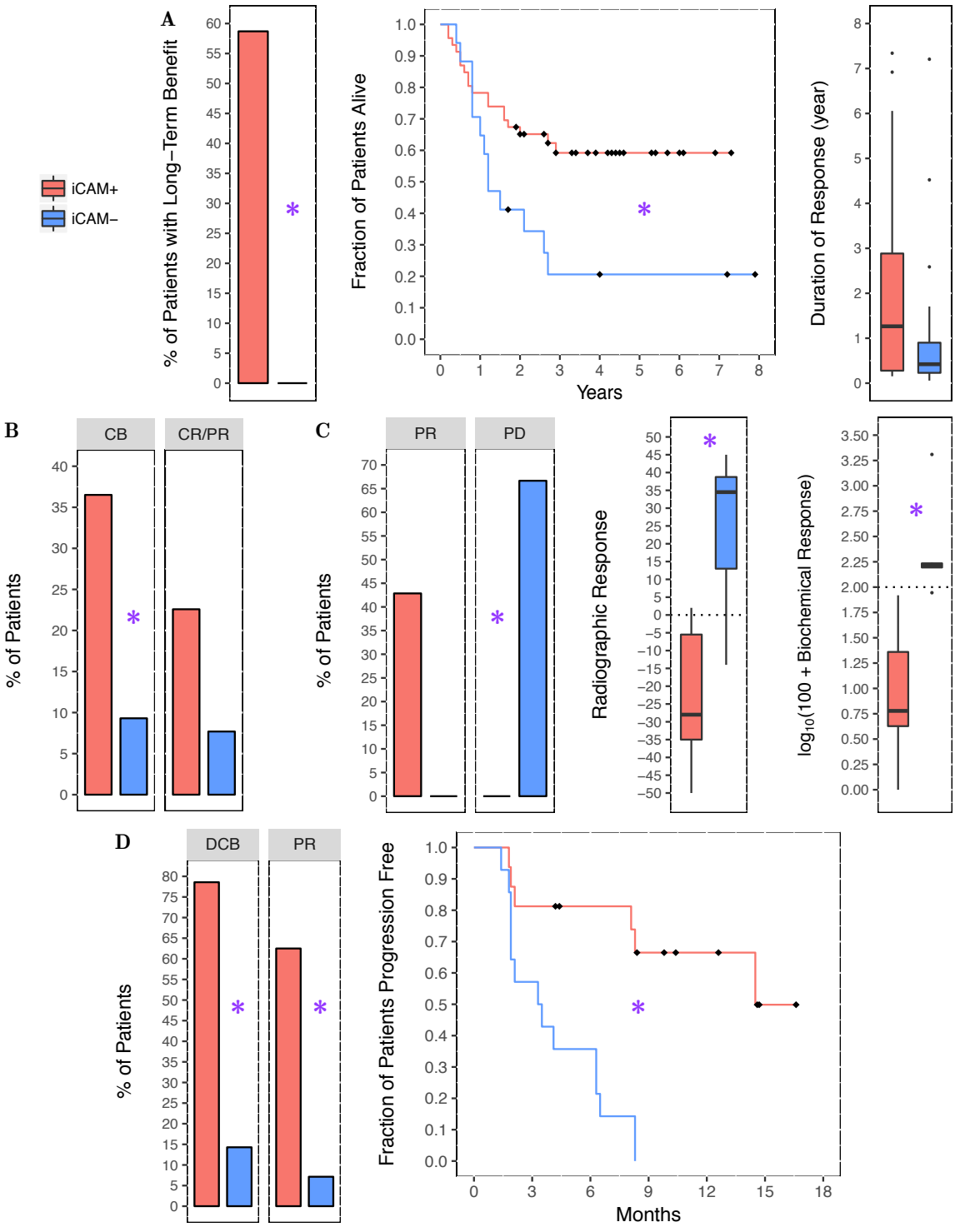


Figure 3.4: iCAM status predicts response to immune checkpoint therapy in published datasets.

Comparison of response by iCAM status: for CTLA-4 blockade in skin melanoma (A[12], B[13]), and PD-1 blockade in colorectal (C[16]) and lung (D[15]) adenocarcinoma. (D)CB: (durable) clinical benefit, CR/PR: complete/partial response, PD: progressive disease. * difference between iCAM+ and iCAM− is statistically significant at $P < 0.05$ in two-sided Fisher's/logrank/ranksum test. This figure was reprinted with permission from our previously published work [2].

Compared to iCAM− patients, iCAM+ patients had significantly longer overall survival (hazard ratio 0.33 [95% CI: 0.14 to 0.78]), and considerably longer duration of response (median 66 weeks vs 22 weeks). In the Van Allen dataset (panel B of Figure 3.4), $> 35\%$ (23/63) of iCAM+ patients had clinical benefit compared to $< 10\%$ (4/43) of iCAM− patients (odds ratio 5.6 [95% CI: 1.8 to 17.7]). Objective (i.e. complete or partial) response rate was also considerably higher (odds ratio 3.5 [95% CI: 0.9 to 13.1]) for iCAM+ patients (14/62) compared to iCAM− patients (3/39).

For PD-1 blockade, in the Le et al colon adenocarcinoma dataset (panel C of Figure 3.4), objective response rate was higher in iCAM+ patients (3/7) compared to iCAM− patients (0/6), and the fraction of patients with progressive disease was significantly lower in iCAM+ patients (0/7) compared to iCAM− patients (4/6). Radiographic and biochemical response were significantly better in iCAM+ patients than iCAM− patients (median -28.0 vs $+34.5$, -94.0 vs $+62.5$ respectively; Positive: the tumor is growing, Negative: the tumor is shrinking). In the Rizvi dataset for lung adenocarcinoma (panel D of Figure 3.4), around 80% (11/14) of iCAM+ patients had durable clinical benefit compared to $< 15\%$ (2/14) of iCAM− patients (odds ratio 22.0 [95% CI: 3.1 to 157.3]), and $> 60\%$ (10/16) of iCAM+ patients had partial response compared to $< 10\%$ (1/14) of iCAM− patients (odds ratio 21.7 [95% CI: 2.2 to 210.1]). Progression free survival was significantly longer for iCAM+ patients compared to iCAM− patients (hazard ratio 0.13 [95% CI: 0.05 to 0.36]).

3.3.3 iCAM+ tumors are identifiable using FoundationOne or StrandAdvantage assays

To test whether iCAM+ tumors can be identified using routine clinical assays that sequence only a few hundred genes instead of the exome (about 20,000 genes), mutation burden of the tumors from TCGA was recomputed by restricting the WXS data to only the genomic regions assayed by two commercially available assays, FoundationOne (315 genes, `https://www.foundationmedicine.com/genomic-testing/foundation-one`) and StrandAdvantage (selected regions of 152 genes, `https://www.strandcenters.com/strand-advantage`). Panel A of Figure 3.5 shows the receiver operating characteristic (ROC) curves for iCAM+ versus iCAM− class prediction of tumors from TCGA using only mutations in the regions sequenced by FoundationOne and StrandAdvantage assays. This

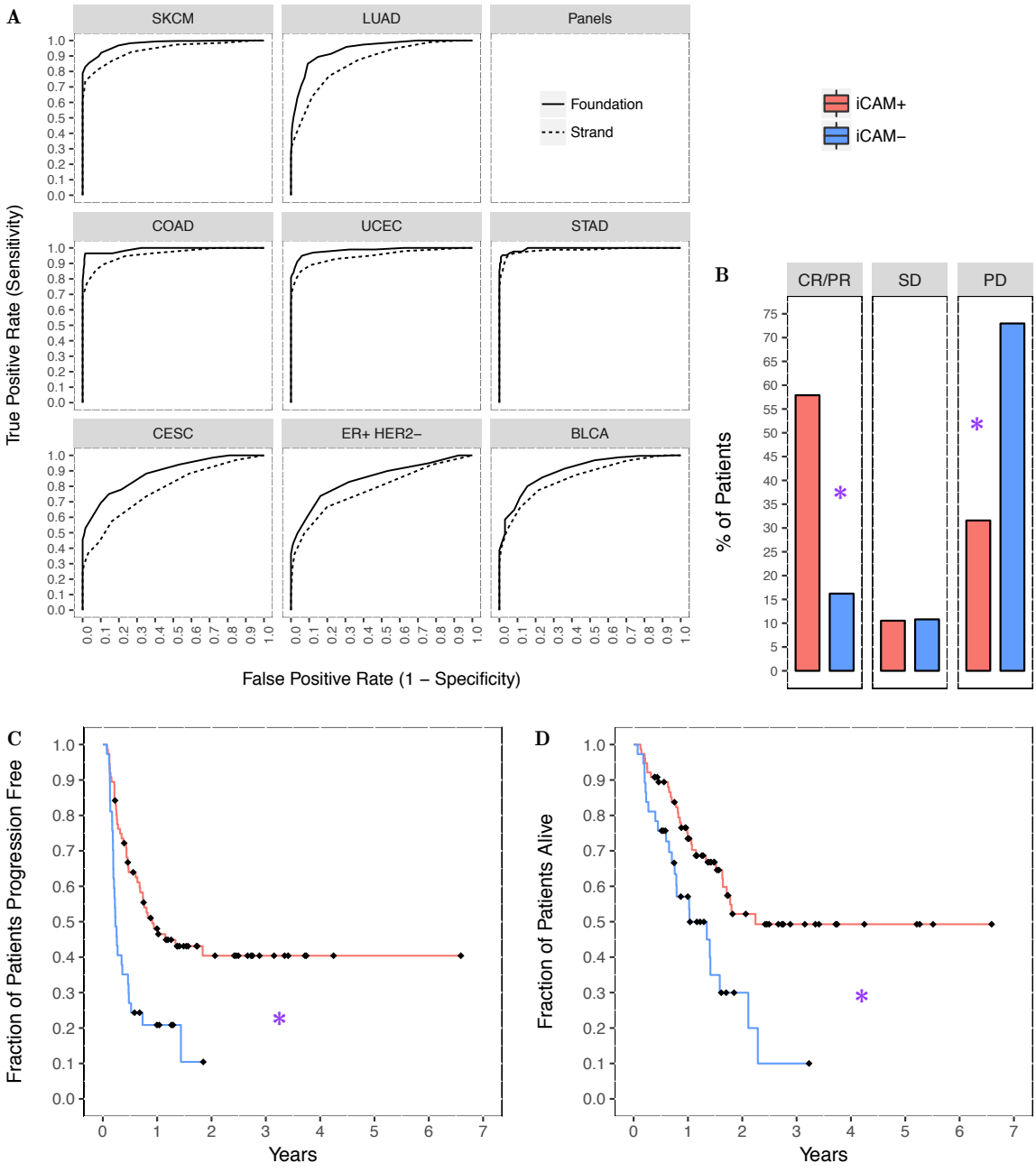


Figure 3.5: Routine clinical sequencing assays can identify iCAM+ tumors with high accuracy.

(A) ROC curves for iCAM status determination of TCGA samples using only the exons assessed by FoundationOne and StrandAdvantage assays. Area under ROC curve: melanoma (0.98, 0.94), lung (0.94, 0.86), colon (0.99, 0.96), endometrial (0.98, 0.95), stomach (0.99, 0.98), cervical (0.88, 0.78), ER+ HER2− breast (0.85, 0.78), and bladder (0.90, 0.86). Comparison of response rates (B), progression free survival (C), and overall survival (D) in 113 patients with metastatic melanoma treated with single agent PD-1 blockade, stratified by iCAM status using the FoundationOne assays. CR/PR: complete/partial response, SD/PD: stable/progressive disease. * difference between iCAM+ and iCAM− is statistically significant at $P < 0.05$ in two-sided Fisher's/logrank test. This figure was reprinted with permission from our previously published work [2].

analysis showed that routine clinical assays like FoundationOne (AUC: 0.85-0.99) and StrandAdvantage (AUC: 0.78-0.98) can identify iCAM+ tumors with high accuracy.

3.3.4 iCAM status predicts response to PD-1 blockade in melanoma

For prospective validation of iCAM status as a biomarker of response to immune checkpoint therapy, mutation and clinical data of 196 de-identified melanoma (111 from R-CINJ, 85 from VICC) were obtained under IRB approved protocols. These samples were sequenced using the FoundationOne assay, which sequenced either 236 or 315 genes. Non-synonymous mutation burden was defined as the number of known or likely mutations + the number of VUS (note: copy number alterations and structural rearrangements did not count towards mutation burden). The iCAM threshold for the FoundationOne assay was estimated to be 9 and 14 non-synonymous mutations respectively for the 236 and 315 genes versions (panel E of Figure 3.2).

113 of the above patients (28 from R-CINJ, 85 from VICC) were treated with single agent PD-1 blockade. Analysis of response data (panel B of Figure 3.5) showed that iCAM+ patients (44/76) had significantly higher (odds ratio 7.10 [95% CI: 2.65 to 19.04]) objective response rate compared to iCAM− patients (6/37). iCAM+ patients also had significantly longer progression free survival (hazard ratio 0.32 [95% CI: 0.18 to 0.57]) and overall survival (hazard ratio 0.37 [95% CI: 0.20 to 0.69]) compared to iCAM− patients, as shown in panel C and D of Figure 3.5. These results are consistent with an earlier observation [14] that mutation load per megabase, as estimated by FoundationOne assay, is predictive of response to PD-1 blockade in a subset (N = 65) of this cohort.

3.3.5 Significantly mutated genes in iCAM+ tumors

The MutSigCV [44] algorithm was used to identify significantly mutated genes in iCAM+ and iCAM− classes in each cancer type at a false discovery rate < 0.1. In colon adenocarcinoma (Figure 3.6), mutations in APC, KRAS, TP53 were frequently observed in iCAM− tumors, consistent with the Vogelstein model of carcinogenesis [47]; but neither APC nor TP53 were significantly mutated in iCAM+ tumors. Instead, iCAM+ tumors had mutations in a set of genes not significantly mutated in

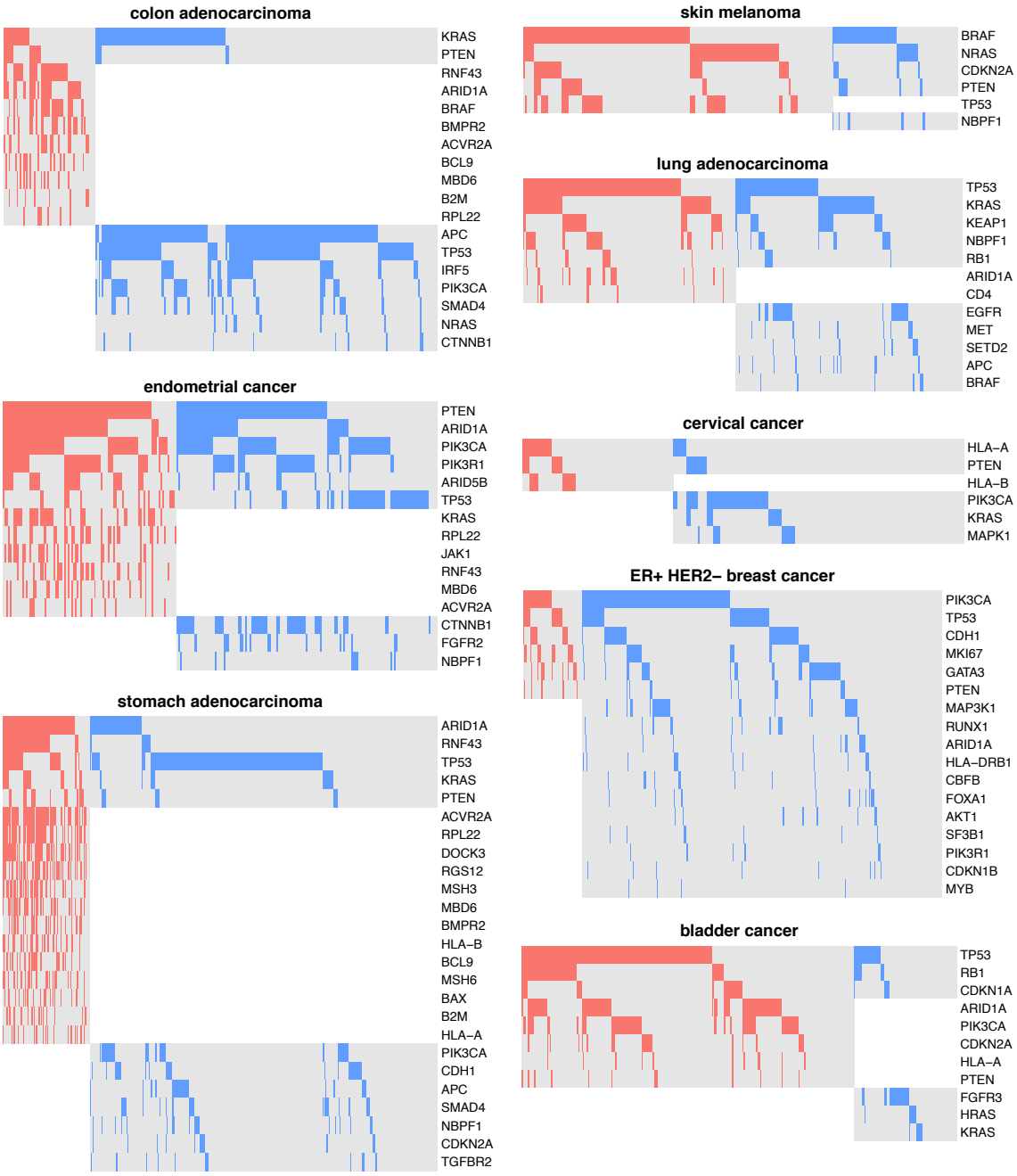


Figure 3.6: iCAM+ and iCAM− tumors of the same tissue have different sets of driver genes.

Columns are tumors, and rows are a subset of significantly mutated genes (MutSigCV [44], false discovery rate < 0.1). White: gene not significantly mutated in that class, gray: gene significantly mutated in that class but not mutated in that particular tumor, red / blue: gene significantly mutated in iCAM+ / iCAM− class and mutated in that particular tumor. This figure was reprinted with permission from our previously published work [2].

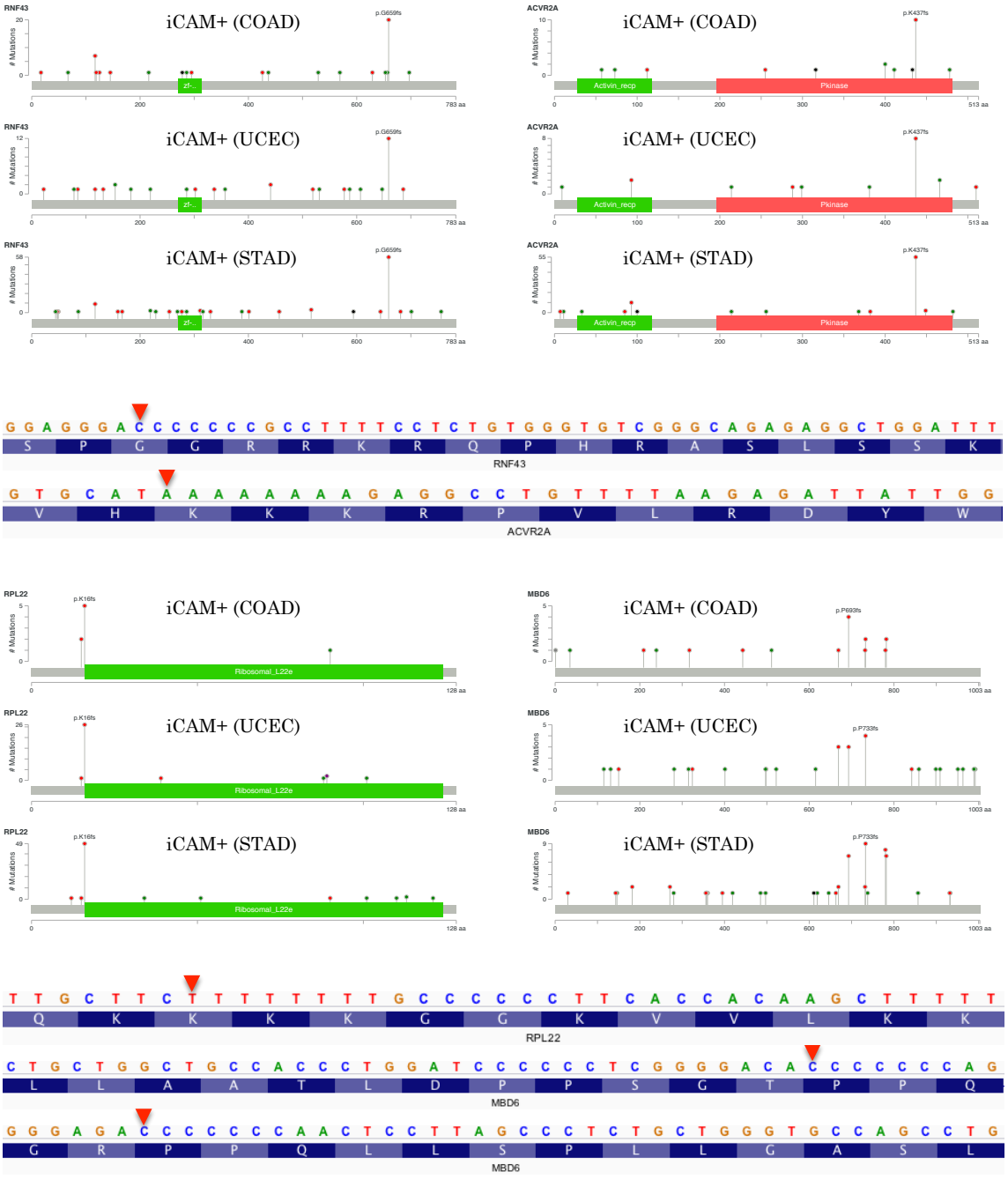


Figure 3.7: Sequence specific hotspots in iCAM+ tumors in colon, endometrial, stomach cancer.

Analysis using cBioPortal [30, 31] showed that iCAM+ tumors in colon adenocarcinoma, endometrial cancer, and stomach adenocarcinoma have mutation hotspots in RNF43, ACVR2A, RPL22, and MBD6. Inspection using IGV [45, 46] showed that the hotspots are located in repeats of length 7-8.

iCAM− tumors, such as RNF43, ACVR2A, RPL22, and MBD6. These 4 genes were also significantly mutated in iCAM+ tumors in endometrial cancer and stomach adenocarcinoma, but rarely mutated in iCAM− tumors (Figure 3.6). Analysis using cBioPortal [30, 31] showed that iCAM+ tumors in colon adenocarcinoma, endometrial cancer, and stomach adenocarcinoma have mutation hotspots in these 4 genes. Inspection of these regions of the genome using IGV [45, 46] showed that all these hotspots are located in repeats of length 7-8. Thus, iCAM+ tumors have sequence specific hotspots in colon adenocarcinoma, endometrial cancer, and stomach adenocarcinoma (Figure 3.7). It should be noted that RNF43 [48], ACVR2A [49], and RPL22 [50] are tumor suppressor genes. In lung adenocarcinoma, EGFR was significantly mutated in iCAM− tumors, but not in iCAM+ tumors. Similarly, several genes were significantly mutated in iCAM+ but not iCAM− tumors or vice versa in the other 4 cancer types. These results suggest that iCAM+ and iCAM− tumors have different sets of driver mutations.

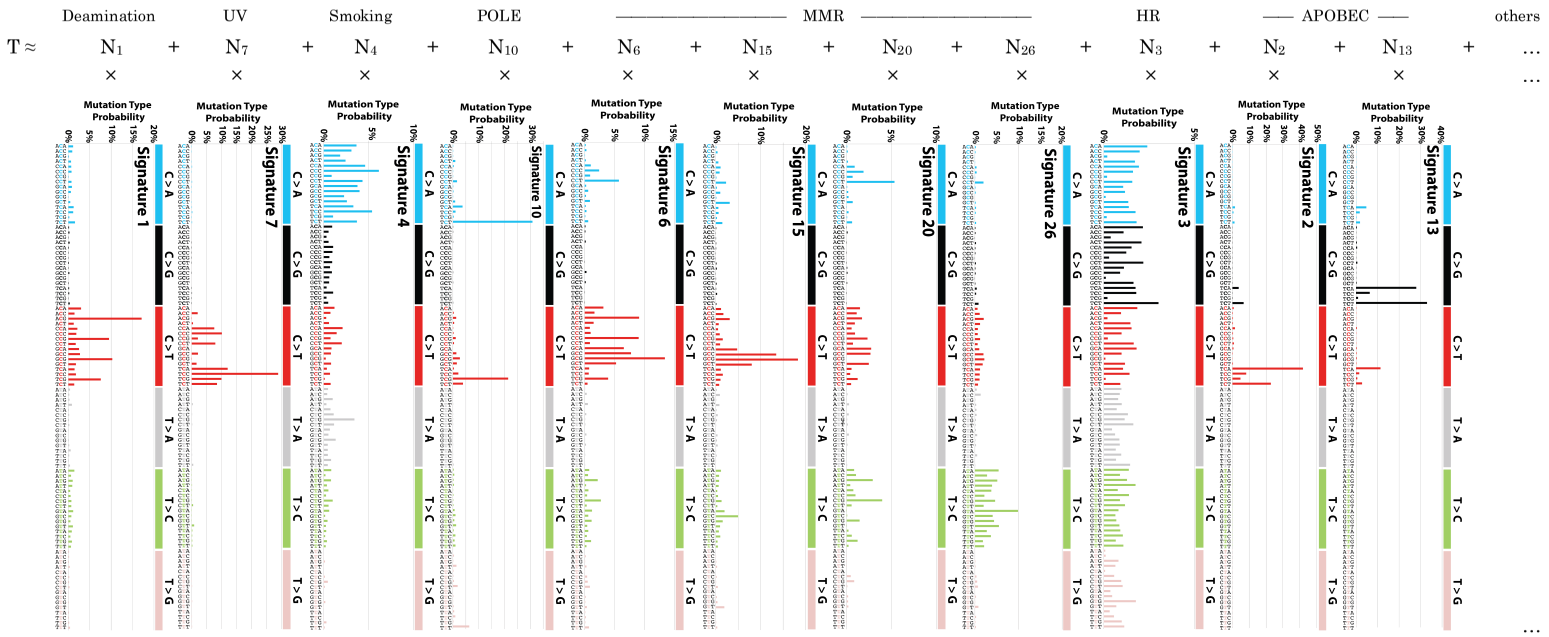
3.3.6 Mutational etiology of iCAM+ tumors

Various etiologies cause mutations in the genome of a cancer cell, but since each etiology has a distinct mutation signature, the mutation pattern present in a tumor can reveal the etiologies that caused that tumor. For each tumor, we estimated the fractional contribution of 30 mutation signatures from COSMIC [21, 22] (`http://cancer.sanger.ac.uk/cosmic/signatures`) using non-negative least squares (Figure 3.8). Analysis of this data showed that iCAM+ tumors had a significantly stronger contribution from UV signature in skin melanoma (SKCM) and smoking signature in lung adenocarcinoma (LUAD) compared to iCAM− tumors (panel A of Figure 3.9), consistent with the known role of exogenous mutagenic exposure in skin [51] and lung [52] carcinogenesis. Molecular signature and history of smoking were also associated with iCAM+ status in lung adenocarcinoma (panel C of Figure 3.9).

In colon adenocarcinoma (COAD), endometrial cancer (UCEC), and stomach adenocarcinoma (STAD), iCAM+ tumors were enriched in either mismatch repair defect or proofreading defect (panel A of Figure 3.9). Consistently, most iCAM+ tumors had either microsatellite instability or POLE

Step 1. use COSMIC convention to represent each tumor as a 96 x 1 "mutation catalogue" T

Step 2. assume each mutation catalogue can be explained using the 30 COSMIC signatures i.e. $T \approx \sum_{i=1}^{30} N_i \times S_i$ for $N_i \geq 0$



Step 3. use non-negative least squares to estimate a 30 x 1 "contribution vector" for each tumor ($T \rightarrow N$)

Step 4. for each tumor, estimate the fractional contribution of various etiologies, as shown below :

$$f_{\text{Deamination}} = N_1 / \sum_{i=1}^{30} N_i \qquad f_{\text{UV}} = N_7 / \sum_{i=1}^{30} N_i \qquad f_{\text{Smoking}} = N_4 / \sum_{i=1}^{30} N_i \qquad f_{\text{POLE}} = N_{10} / \sum_{i=1}^{30} N_i \qquad f_{\text{HR}} = N_3 / \sum_{i=1}^{30} N_i$$

$$f_{\text{MMR}} = (N_6 + N_{15} + N_{20} + N_{26}) / \sum_{i=1}^{30} N_i \qquad f_{\text{APOBEC}} = (N_2 + N_{13}) / \sum_{i=1}^{30} N_i \qquad f_{\text{others}} = 1 - (f_{\text{Deamination}} + f_{\text{UV}} + f_{\text{Smoking}} + f_{\text{POLE}} + f_{\text{MMR}} + f_{\text{HR}} + f_{\text{APOBEC}})$$

Figure 3.8: Protocol for estimating fractional contribution of mutational etiologies in each tumor.

This figure was reprinted with permission from our previously published work [2]. The depicted mutation signatures are from COSMIC. (`http://cancer.sanger.ac.uk/cosmic/signatures`)

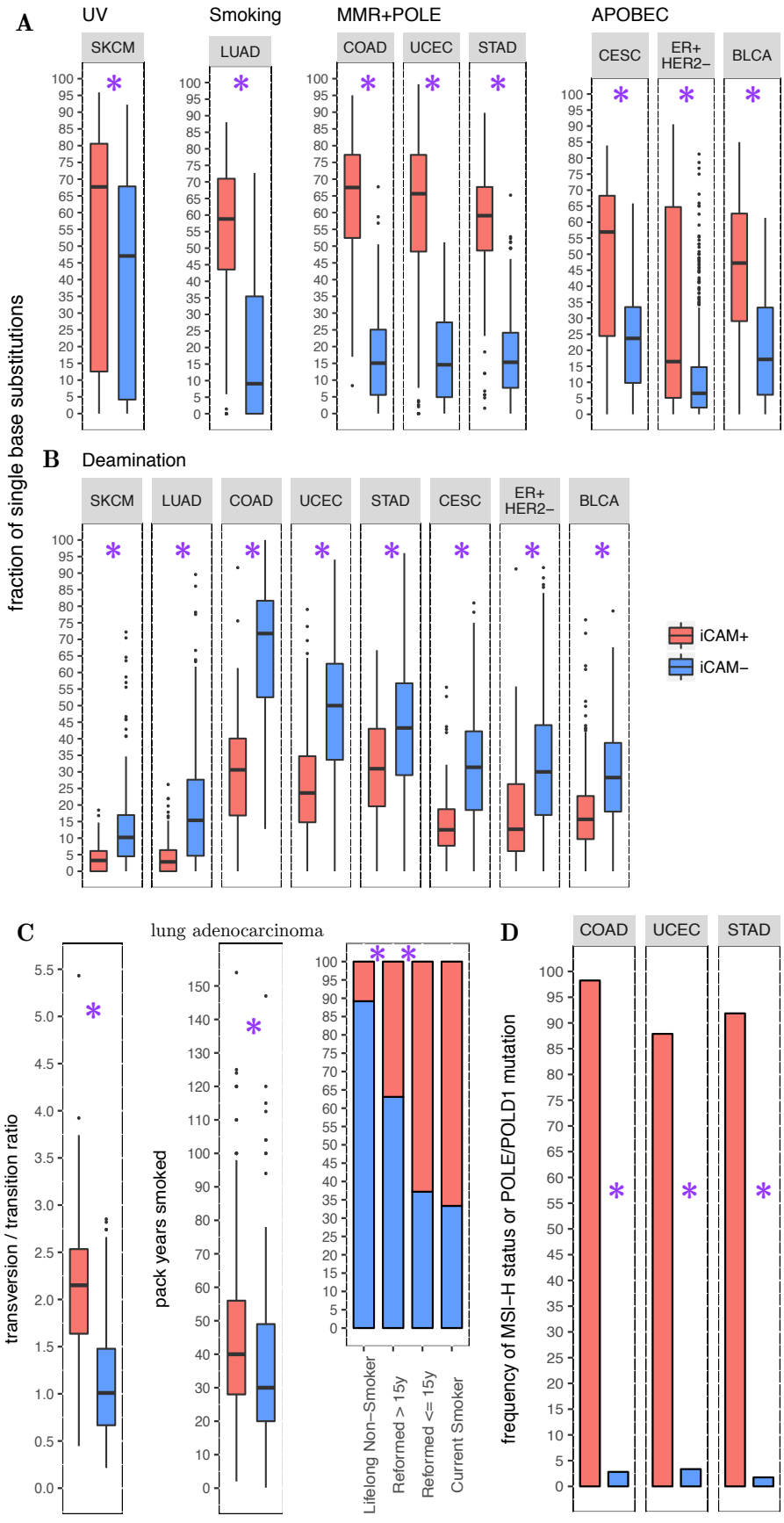


Figure 3.9: iCAM+ and iCAM− tumors of the same tissue have different mutational etiology.

(A) iCAM+ tumors have a stronger contribution from mutation signatures associated with UV in melanoma and smoking in lung adenocarcinoma; proofreading defect or mismatch repair defect in colon, endometrial, and stomach cancer; and aberrant APOBEC activity in cervical, ER+ HER2− breast, and bladder cancer. (B) iCAM− tumors have a stronger contribution from deamination of 5-methyl cytosine (associated with aging) in all 8 cancer types. Consistently, iCAM+ status was associated with (C) molecular signature of smoking, as well as smoking history in lung adenocarcinoma, and (D) POLE/POLD1 mutation or MSI-H status in colon, endometrial, and stomach cancer. * difference between iCAM+ and iCAM− is statistically significant at $P < 0.05$ in two-sided ranksum / Fisher's test. This figure was reprinted with permission from our previously published work [2].

mutation in colon adenocarcinoma, endometrial cancer, and stomach adenocarcinoma (panel D of Figure 3.9). It should be noted that mismatch repair defect was known to be associated with response to PD-1 blockade in colorectal cancer [16], and was subsequently shown (in an independent cohort) to be associated with response to PD-1 blockade in other solid cancer types [53]. Proofreading defect may also be associated with response to PD-1 blockade, as discussed in details in Chapter 2.

In cervical cancer, ER+ HER2− breast cancer, and bladder urothelial cancer, iCAM+ status was associated with APOBEC mutation signatures (panel A of Figure 3.9). The presence of APOBEC mutagenesis has recently been reported in these cancer types [54]. Of note, in ER+ HER2− breast cancer, around 25% of Luminal-B tumors were iCAM+, compared to $< 10\%$ of Luminal-A tumors. This suggests that Luminal-B tumors, which have a relatively poor prognosis under conventional treatment [55], are more likely to respond to immune checkpoint therapy compared to Luminal-A tumors.

In all 8 cancer types, iCAM− tumors had a significantly stronger contribution from deamination of 5-methyl cytosine, indicative of aging [23], compared to iCAM+ tumors (panel B of Figure 3.9).

3.3.7 Immune micro-environment of iCAM+ tumors

In most of the 8 cancer types, natural killer cells constituted a significantly higher fraction of leukocytes, M1 macrophages constituted a significantly higher fraction of macrophages, and regulatory T cells constituted a significantly lower fraction of T cells, in iCAM+ tumors compared to iCAM− tumors (Figure 3.10). Since natural killer cells have important anti-tumor activity [38], M1 macrophages curtail tumor growth by suppressing angiogenesis and inducing apoptosis [39], and regulatory T cells suppress immune response [56], this shows that iCAM+ tumors have a more favorable immune microenvironment.

61 genes were significantly over-expressed in iCAM+ tumors compared to iCAM− tumors in all 8 cancer types, and about 70% of these genes were immune system related, including genes associated with the interferon-gamma pathway (IFNG, CXCL9, CXCL10, CXCL11) and CD8+ T cells (CD8A, PRF1, GZMA, GZMB). Thus, immune response was the most important difference

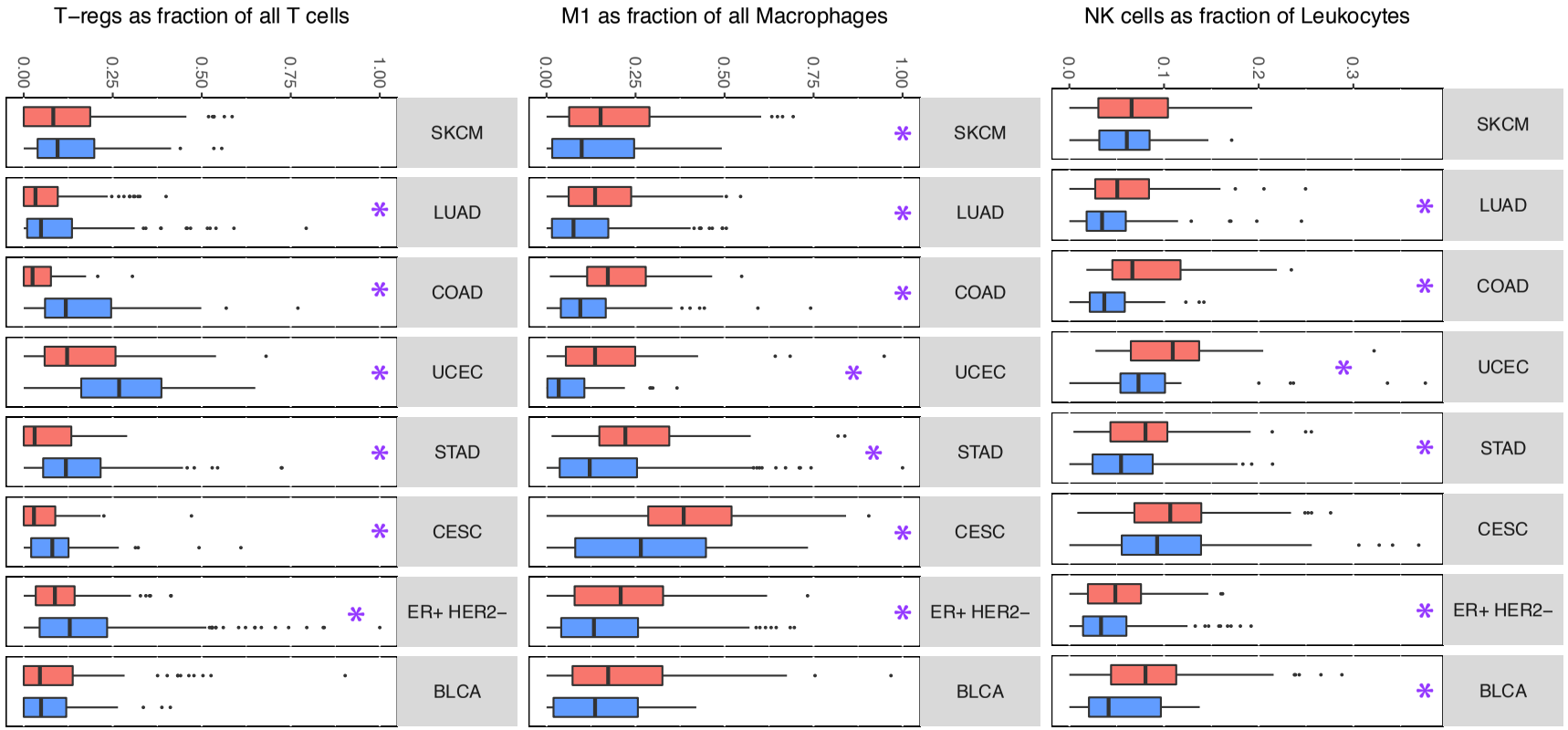


Figure 3.10: iCAM+ and iCAM− tumors of the same tissue have different micro-environment.

Comparison of iCAM+ and iCAM− tumors for Natural Killer (NK) cells as fraction of leukocytes, M1 polarized macrophages as fraction of all macrophages, and regulatory T cells (T-regs) as fraction of all T cells. * difference between iCAM+ and iCAM− is statistically significant at $P < 0.05$ in two-sided Wilcoxon ranksum test. This figure was reprinted with permission from our previously published work [2].

between iCAM+ and iCAM− tumors. The immune checkpoint gene LAG-3 was significantly over-expressed in iCAM+ tumors in all 8 cancers types, and may also be a good target for immune checkpoint therapy.

3.4 Discussion

In summary, in 8 out of 33 solid cancer types in TCGA, a mutation burden threshold (iCAM) can identify tumors with signatures of immune checkpoint activation. Analysis of published datasets in melanoma, lung adenocarcinoma and colon cancer showed that patients with iCAM+ tumors have better response to immune checkpoint therapy. iCAM+ tumors can be identified using clinical sequencing assays, and iCAM status determined using the FoundationOne assay can predict response to single agent PD-1 blockade in melanoma. Although the FoundationOne and StrandAdvantage assays were used in this study, any hybrid-capture based assay that identifies mutations in a substantial portion of the exome could be used instead. The thresholds we found could be further refined using higher sequencing depth, and detailed analysis of tumor purity and mutant allele frequency.

Although a clinically significant subset of patients respond to immune checkpoint therapy in several other cancer types, such as triple negative breast cancer [57] and clear-cell renal cancer [58], an iCAM threshold could not be identified in these cancer types. This indicates that some other mechanisms of immune activation may be operative in those cancer types. Furthermore, not all iCAM+ patients responded to PD-1 or CTLA-4 blockade in our 8 cancer types, possibly because here too, other mechanisms of immune-evasion may be operative, such as mutations in class I HLA genes [59] or abundant MDSC [60]. Similarly, some iCAM− patients did respond to PD-1 or CTLA-4 blockade, possibly because the immune response was triggered by mechanisms other than a high mutation burden (e.g. expression of exogenous viruses [18, 19, 20], expression of endogenous retrovirus due to hypo-methylation [61, 62]). Low tumor purity may also have affected the ability to detect mutations accurately. Hence, the iCAM status should only be considered as one of several potential biomarkers of response to immune checkpoint therapy.

Chapter 4

Expression of exogenous virus in tumor as a biomarker

Published as: Panda A, Mehnert JM, Hirshfield KM, et al. J Natl Cancer Inst. 2018; 110(3) [3]

4.1 Introduction

Hyper-mutation in tumor due to a proofreading defect, or some other etiologies, is a potential biomarker of response to immune checkpoint therapy in several solid cancer types, as we have noted in previous chapters. However, hyper-mutation can not be the only biomarker of response to immune checkpoint therapy, since tumors with low mutation burden, such as in MCPyV+ Merkel-cell cancers, have recently been shown to respond to PD-1 blockade [18, 19]. Since one of the functions of the immune system is to identify cells infected by pathogens, it is plausible that there is immune recognition of tumors that express viral element. If this is true, some virally mediated tumors should also trigger an anti-tumor immune response spearheaded by CD8+ T cells, that the tumors may block using immune checkpoint pathways, which again may make the tumors vulnerable to immune checkpoint therapy. Motivated by this hypothesis, we analyzed the following virally mediated tumors: EBV+ gastric cancer, HPV+ cervical and head-neck squamous-cell cancer, and HBV+ or HCV+ (hepatitis B/C virus) liver cancer for evidence of immune checkpoint activation.

4.2 Results from TCGA gastric cancer cohort

4.2.1 Study design

A subset of advanced gastric cancer is known to respond to PD-1 blockade [63], and EBV-associated NK/T cell lymphoma has recently been shown to respond to PD-1 blockade [20], which suggests the

possibility that EBV+ gastric cancer may also respond to immune checkpoint therapy. Notably, a substantial subset of gastric cancers have microsatellite instability (as noted in Chapter 3), and may also be sensitive to immune checkpoint therapy by virtue of hyper-mutation due to a mismatch repair defect [53]. Since EBV+ and MSI status were mutually exclusive in the TCGA gastric cancer cohort, we classified the TCGA gastric cancer samples into 3 groups: (a) known EBV+ tumors (EBV), (b) tumors with known microsatellite instability (MSI), and (c) the rest of the tumors (MSS).

4.2.2 Mutation Burden

Analysis of WXS based somatic mutation data (panel A of Figure 4.1) showed that EBV+ tumors (N = 25) had a non-synonymous mutation burden comparable to that of MSS tumors (N = 277), an order of magnitude lower than the non-synonymous mutation burden of MSI tumors (N = 69). Thus, EBV status may be a predictor of immune checkpoint activation in iCAM− gastric cancer.

4.2.3 Immune Activation

In spite of a low mutation burden, EBV+ gastric tumors (N = 25) had significantly higher immune infiltration (ImmuneScore) and mRNA expression of CD8A (a marker of CD8+ T cell), not just compared to MSS tumors (N = 310), but even compared to MSI tumors (N = 80), suggesting the presence of immune activation spearheaded by CD8+ T cells (panel B of Figure 4.1).

For a more comprehensive evaluation, we compared the expression of about 700 immune-related genes among the 3 groups pairwise, and found that approximately two thirds of these genes were differentially expressed, at 1% false discovery rate (Benjamini Hochberg). The heatmap (panel C) in Figure 4.1 shows the relative expression of these differentially expressed immune-related genes (rows) in the tumors (column), where the top quartile is colored gold and the bottom half is colored cyan. This demonstrated that EBV+ tumors (red) had high expression of a large fraction of immune-related genes compared to MSS tumors (blue), whereas MSI tumors (green) had an intermediate phenotype.

As a consistency check, high resolution digital images of H&E stained FFPE sections of 17

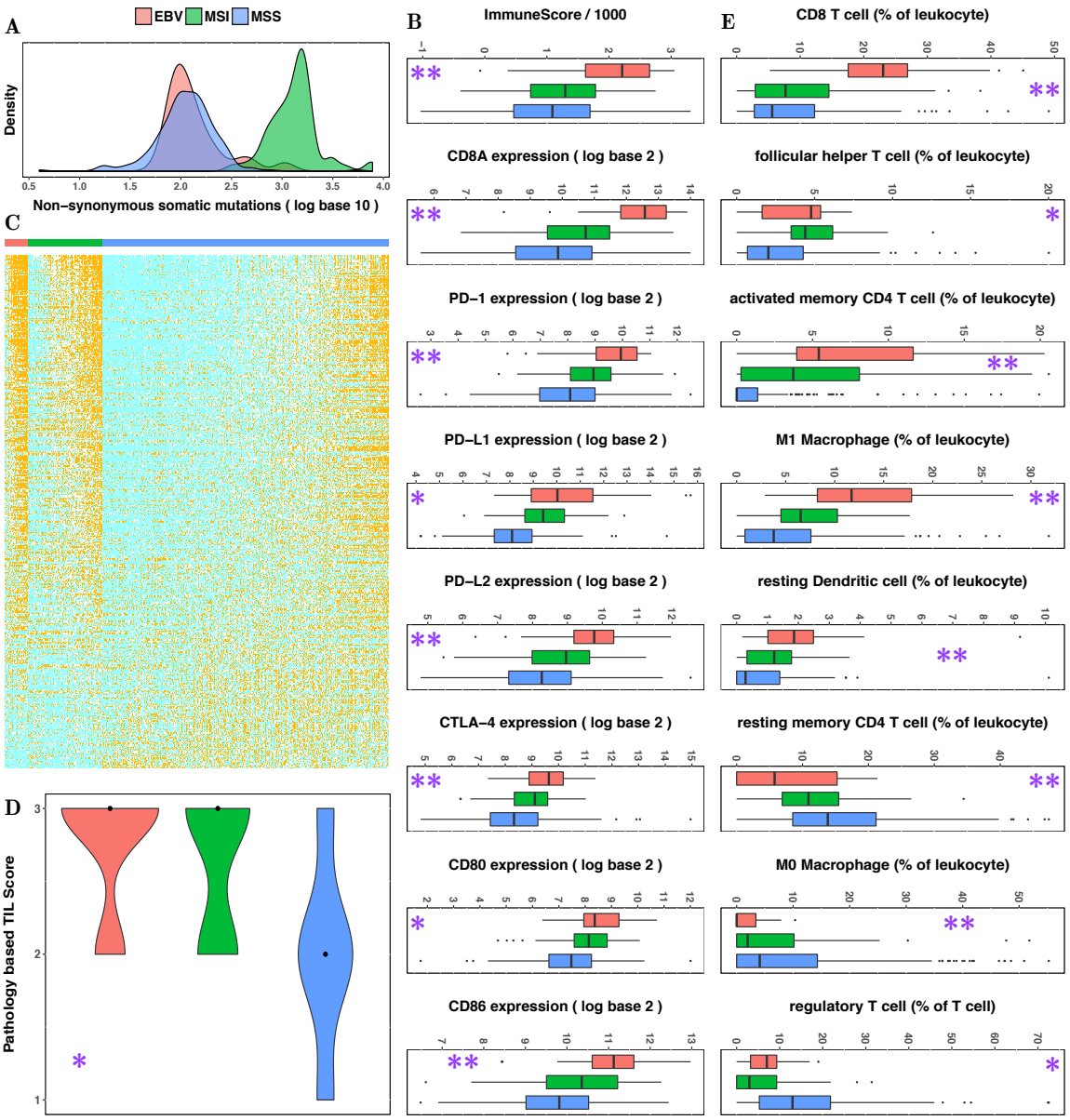


Figure 4.1: Comparison of EBV+, MSI, and MSS tumors in TCGA gastric cancer cohort.

(A) Number of non-synonymous mutations. (B) Overall immune infiltration, mRNA level of CD8+ T cell marker CD8A, and immune checkpoint genes of PD-1/CTLA-4 pathways. (C) Heatmap of relative expression of differentially expressed immune-related genes (gold: high, cyan: low). (D) Pathology-based lymphocyte infiltration scores (in 17 EBV+, 13 MSI, 16 MSS tumors; 1: low, 3: high). (E) Fractional composition of tumor-infiltrating leukocytes, and proportion of regulatory T cells (T-regs) among all T cells. ** both EBV+ vs MSI and EBV+ vs MSS differences are statistically significant, * EBV+ vs MSS difference is statistically significant, at $P < 0.05$ in two-sided Wilcoxon ranksum test. This figure was reproduced from our previously published work [3].

EBV+, 13 MSI, 16 MSS tumors from TCGA gastric cancer cohort were evaluated by a pathologist who was blinded to the results from genomic analysis. The pathologist scored each image on a scale of 1 (minimal) to 3 (robust) for the presence of tumor infiltrating lymphocytes. The distribution of this pathology based lymphocyte infiltration score in the 3 groups (panel D of Figure 4.1) showed that EBV+ tumors had a significantly higher lymphocytic infiltration compared to MSS tumors, whereas MSI tumors had an intermediate phenotype, consistent with the findings from RNA-seq data. This was also consistent with the findings of 2 pathological studies that reported lymphocyte infiltration in 2 independent cohorts of EBV+ gastric cancer [64, 65]. In summary, these results confirmed the presence of immune activation in EBV+ gastric cancer.

4.2.4 Checkpoint pathway up-regulation

EBV+ tumors (N = 25) had a significantly higher mRNA expression of the checkpoint receptors PD-1 and CTLA-4, and their ligands (PD-L1, PD-L2; CD80, CD86) compared to MSS tumors (N = 310); and significantly higher mRNA expression of the checkpoint receptors PD-1 and CTLA-4, and their ligands PD-L2 and CD86 (but not the ligands PD-L1 and CD80) compared to MSI tumors (N = 80), as shown in panel B of Figure 4.1. These results confirmed the up-regulation of both PD-1 and CTLA-4 immune checkpoint pathways in EBV+ gastric cancer. These results are also consistent with the findings of 3 recent immunohistochemistry studies [66, 67, 68] that reported PD-L1 expression in a large fraction of EBV+ gastric cancer in 3 independent cohorts.

4.2.5 Immune microenvironment

As shown in panel E of Figure 4.1, CD8+ T cells and resting dendritic cells constituted a significantly higher fraction of tumor infiltrating leukocytes in EBV+ tumors (N = 23) compared to both MSI tumors (N = 58) and MSS tumors (N = 189). The relative abundance of activated memory CD4 T cells and M1 macrophages were also significantly higher, and the relative abundance of resting memory CD4 T cells and M0 macrophages were significantly lower, in EBV+ tumors compared to both MSI tumors and MSS tumors (panel E of Figure 4.1). This suggests M1 polarization of

macrophages and activation of memory CD4 T cells in EBV+ tumors. Additionally, compared to MSS tumors, follicular helper T cells constituted a significantly higher fraction of leukocytes, and regulatory T cells constituted a significantly lower fraction of T cells in EBV+ tumors. Since M1 macrophages can curtail tumor growth [39], and regulatory T cells suppress immune response [56], this suggests that EBV+ tumors have a more favorable immune microenvironment; i.e. they are primed to respond if the block on the immune response is removed, as is the case with immune checkpoint therapy.

4.2.6 Prediction

To summarize, as we describe in [3], analysis of the TCGA gastric cancer cohort showed that in spite of low non-synonymous mutation burden, EBV+ gastric cancers showed strong evidence of immune activation (led by CD8+ T cells) and checkpoint (both PD-1 and CTLA-4) pathway up-regulation. Also, although EBV+ and MSI tumors represented 2 distinct mechanisms of immune activation (via expression of exogenous virus in tumor and hyper-mutation in tumor respectively), MSI tumors had an intermediate phenotype for almost all immunological factors. These results suggest that EBV+ gastric cancer, a subset of low mutation burden gastric cancer, may respond to immune checkpoint therapy even better than MSI gastric cancer.

4.3 Results from other virally mediated cancers

Motivated by the evidence of EBV associated immune checkpoint activation in gastric cancer, we subsequently analyzed other virally mediated tumors from TCGA. We also extended the analysis beyond PD-1 and CTLA-4 pathways, to include other genes in immune checkpoint pathways [10].

Analysis of TCGA cervical cancer (CESC) dataset (Figure 4.2) showed that HPV− tumors had significantly lower immune infiltration (ImmuneScore) and mRNA expression of CD8A (a marker of CD8+ T cells) compared to HPV+ tumors, suggesting the absence of immune activation led by CD8+ T cells in HPV− cervical cancer. mRNA expression of checkpoint receptors PD-1 and CTLA-4, their ligands (PD-L1, PD-L2; CD80, CD86), as well as other genes of immune checkpoint pathways

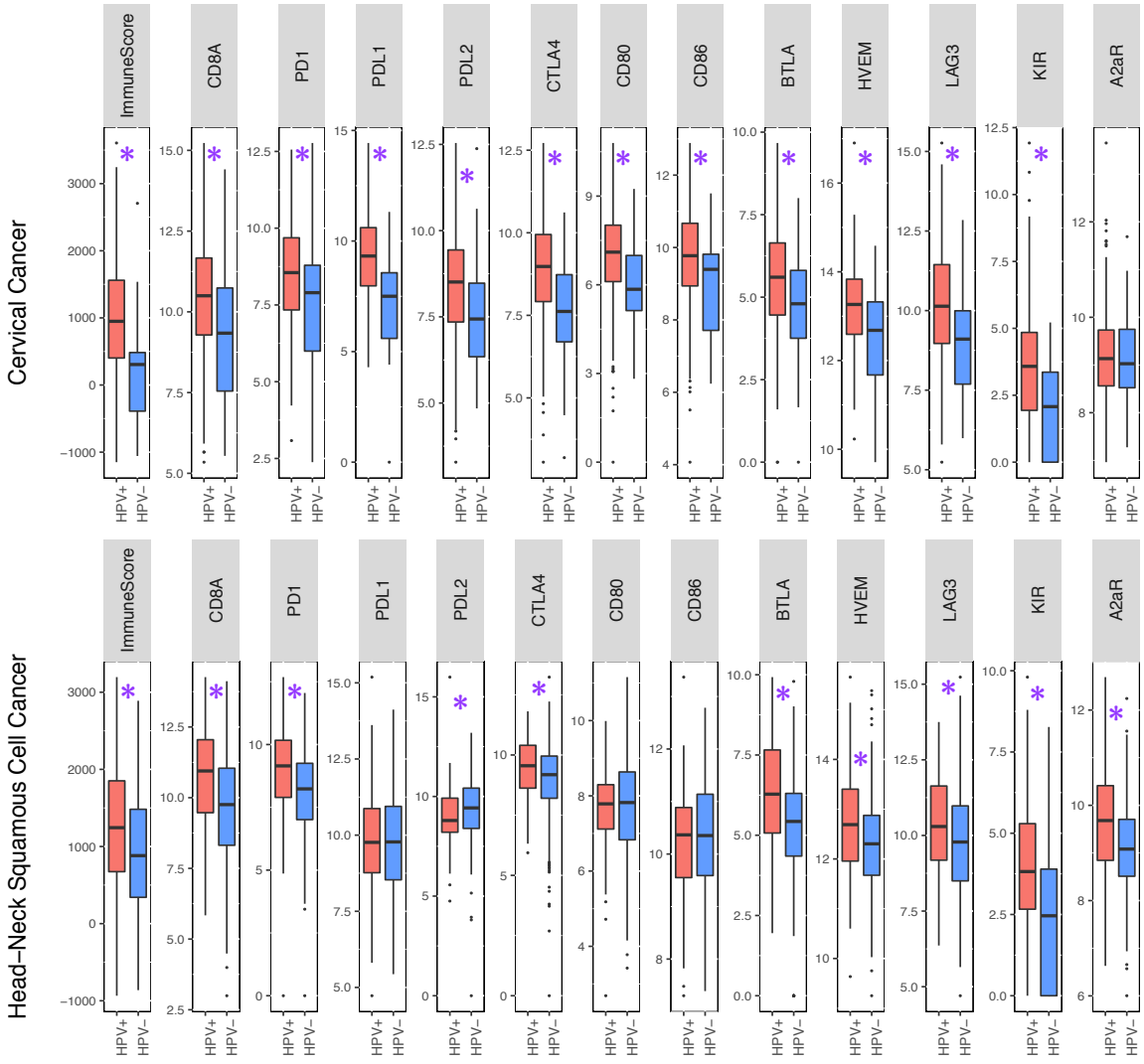


Figure 4.2: Immunological comparison between HPV+ and HPV− tumors in TCGA dataset.

* HPV+ vs HPV− difference is statistically significant at $P < 0.05$ in two-sided ranksum test.

(e.g. BTLA and HVEM, LAG-3, KIR) were significantly lower in HPV− tumors compared to HPV+ tumors, suggesting the absence of checkpoint pathway up-regulation in HPV− cervical cancer. These results suggest that HPV− cervical cancers are unlikely to respond to immune checkpoint therapy.

Analysis of TCGA head-neck squamous cell cancer (HNSC) dataset (Figure 4.2) showed that HPV+ tumors had significantly higher immune infiltration (ImmuneScore) and mRNA expression of CD8A (a marker of CD8+ T cells) compared to HPV− tumors, suggesting the presence of immune activation led by CD8+ T cells in HPV+ HNSC. Although mRNA expression of checkpoint receptors PD-1 and CTLA-4 were significantly higher, mRNA expression of their ligands were NOT significantly higher in HPV+ tumors compared to HPV− tumors. Hence HPV status alone is unlikely to predict response to PD-1 blockade or CTLA-4 blockade in HNSC. However, mRNA expression of other immune checkpoint genes, such as BTLA and HVEM, LAG-3, KIR, A2aR were significantly higher in HPV+ tumors compared to HPV− tumors in HNSC. Thus, in HNSC, HPV status may be a biomarker of response to immune checkpoint therapies other than PD-1 or CTLA-4 blockade (e.g. treatment with the novel drug that target LAG-3).

Finally, analysis of the TCGA liver hepatocellular cancer cohort showed that neither HBV+ nor HCV+ status was associated with immune activation and checkpoint pathway up-regulation (data not shown). Hence other mechanism(s) of immune checkpoint activation is operative in liver hepatocellular cancers that respond to immune checkpoint therapy. This negative result also shows that not all virally mediated tumors are immunogenic.

4.4 Validation in an EBV+ gastric cancer patient

4.4.1 Response to PD-1 blockade in a gastric cancer patient at R-CINJ

A 53-year-old patient [3] was diagnosed with Stage IIIC gastric cancer, and underwent chemotherapy, followed by surgery, followed by radiotherapy and chemotherapy. She had local and distant disease recurrence within 16 months after surgery, and received more chemotherapy, but was hospitalized 15 months later after her condition worsened. She was then enrolled in an IRB-approved Phase I trial of PD-1 blockade (avelumab at 10 mg/kg every 2 weeks) at R-CINJ with informed consent,

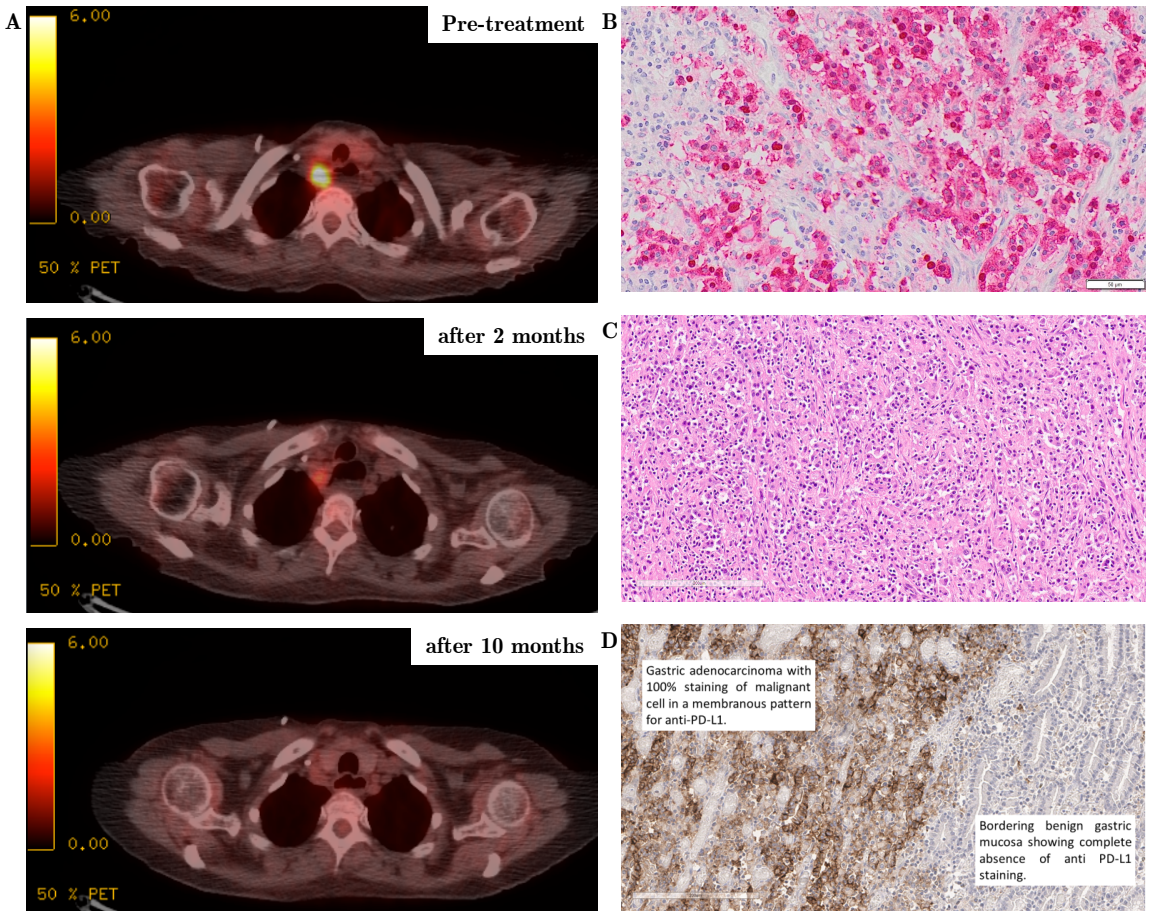


Figure 4.3: Clinical benefit from PD-1 blockade in a patient with metastatic EBV+ gastric cancer.

(A) PET-CT images taken before and after (2, 10 months) treatment. (B) Staining of primary tumor for EBV-encoded RNA (EBER) in red showed that the tumor is EBV+ (internal negative control: normal gastric mucosa). (C) Gastric biopsy showed intense lymphocyte infiltration in the tumor. (D) Staining of gastric biopsy for PD-L1 expression showed that the tumor cells (in left) are strongly positive for PD-L1, but the benign gastric mucosa (in right) is negative for PD-L1. This figure was reproduced from our previously published work [3].

and had remarkable clinical and radiographic response to PD-1 blockade (panel A of Figure 4.3).

4.4.2 Staining for EBV-encoded RNA confirmed that the tumor is EBV+

With informed consent, her primary tumor was sequenced using FoundationOne assay. The result showed a PIK3CA hotspot mutation (E545K), an ARID1A frameshift mutation (N1203fs*3), PTEN loss, and no HER2 amplification. It should be noted that co-mutation of PIK3CA and ARID1A, as observed in this patient's tumor, is frequently observed in EBV+ gastric adenocarcinoma [29]. The mutation burden of the tumor was low, and not consistent with the presence of a mismatch repair or proofreading defect in the tumor. Therefore, I suggested that the tumor be tested for EBV, which was carried out. Staining for EBV-encoded RNA (EBER) showed that the tumor was strongly positive for EBV, as shown in panel B of Figure 4.3.

4.4.3 Pathology confirmed lymphocyte infiltration and PD-L1 overexpression

Histologic assessment of the tumor showed an abundance of infiltrating lymphocytes in the tumor (panel C of Figure 4.3), and strong expression of PD-L1 in tumor cells (panel D of Figure 4.3). Tumor infiltrating lymphocytes in the same sample also showed expression of PD-1 (not shown). These observations supported our results from TCGA dataset that EBV+ gastric cancers show evidence of immune activation and checkpoint pathway upregulation.

4.4.4 Study conclusions

The remarkable clinical benefit observed in this patient with metastatic EBV+ gastric cancer, and the histologic evidence of immune activation and PD-1 pathway up-regulation in this patient's tumor, support the prediction that, in spite of a low mutation burden, EBV+ gastric cancers may be sensitive to immune checkpoint therapy. EBV+ tumors are frequently found also in nasopharyngeal cancer and Hodgkin's/Burkitt's lymphoma, and it should be investigated whether EBV+ tumors of these cancer types are also likely to respond to immune checkpoint therapy. Recent data showing response to PD-1 blockade in EBV-associated NK/T cell lymphoma also support this concept [20].

Chapter 5

Expression of endogenous retrovirus in tumor as a biomarker

5.1 Introduction

As discussed in the previous chapter, expression of certain exogenous viruses in tumors is a potential biomarker of response to immune checkpoint therapy in some cancer types. This suggests the possibility that expression of certain endogenous retroviruses (hERVs) in tumors may also be associated with response to immune checkpoint therapy in some solid cancer types.

Endogenous retroviruses are viral elements that constitute 8% of human genome [69], but are not expressed in normal somatic tissues (for comparison, protein coding regions constitute only 1% of the human genome [70]). They are also abundant in the genome of non-human mammals, and it is not entirely clear whether they are the remnants of ancient retroviral infections or are precursors of exogenous retroviruses. Epigenetic alterations in tumor may lead to expression of hERVs in tumors, but since these hERVs are usually not expressed in normal somatic tissues, the immune system may be able to recognize tumor cells expressing certain hERVs as non-self. If so, tumors expressing these immunogenic hERVs may trigger an anti-tumor immune response led by CD8+ T cells, that the tumors may block using immune checkpoint pathways, which makes these tumors vulnerable to immune checkpoint therapy.

Motivated by this hypothesis, we analyzed the TCGA dataset to search for evidence of hERV associated immune checkpoint activation in solid tumors. A recent study [24] quantified the expression of 66 transcribed [71] hERVs in thousands of tumors from TCGA by re-mapping the raw RNA-seq data of these tumors. In this dataset, both mRNA expression data and hERV expression data were available for 4,910 tumors of 21 solid cancer types, whose analysis will be the focus of this chapter.

5.2 Results

5.2.1 Identification of potentially immunogenic hERVs in solid cancer types

In each cancer type, we looked for hERVs whose expressions were significantly correlated with immune activation (immune infiltration in tumor, and expression of CD8A) and expression of checkpoint (PD-1/CTLA-4) pathway genes. Such potentially immunogenic hERVs were frequent in 3 cancer types: 20 in clear-cell renal cancer (panel A of Figure 5.1), 12 in ER+ HER2− breast cancer (panel B of Figure 5.1), and 9 in colon adenocarcinoma (panel C of Figure 5.1). Although some potentially immunogenic hERVs in colon cancer (3/9) were from the hERV-H family, most of the potentially immunogenic hERVs were from the hERV-K family (18/20, 10/12, and 4/9 in clear-cell renal cancer, ER+ HER2− breast cancer, and colon adenocarcinoma respectively). This is probably because hERV-K is the most recently integrated [72] and best-preserved [72] family of hERV that has retained a non-defective gag gene [73] and ORFs in pol and env genes [73], and its simian equivalent is known to induce an immune response in Indian rhesus macaque [74]. It should also be noted that expression of some members of the hERV-K family has been reported [24] to be correlated with expression of cytotoxins granzyme-A and perforin in clear-cell renal cancer and breast cancer.

5.2.2 Classification of tumors by expression of potentially immunogenic hERVs

Hierarchical clustering of tumors using expression (percentile) of potentially immunogenic hERVs stratified the tumors into 3 groups in clear-cell renal cancer (panel D of Figure 5.1), ER+ HER2− breast cancer (panel E of Figure 5.1), and colon cancer (panel F of Figure 5.1): tumors with high/ intermediate/low expression of potentially immunogenic hERVs. As mentioned above, in clear-cell renal cancer and ER+ HER2− breast cancer, potentially immunogenic hERVs were mostly members of the hERV-K family, and these members were strongly co-expressed, making the 3 clusters visually obvious (panel DE of Figure 5.1). In colon cancer however, potentially immunogenic hERVs included members of both hERV-K and hERV-H family, and members of hERV-K and hERV-H family were weakly co-expressed making the 3 clusters less obvious (panel F of Figure 5.1). This is consistent with a recent report [75] that hERV-K and hERV-H expression are weakly correlated in colon cancer.

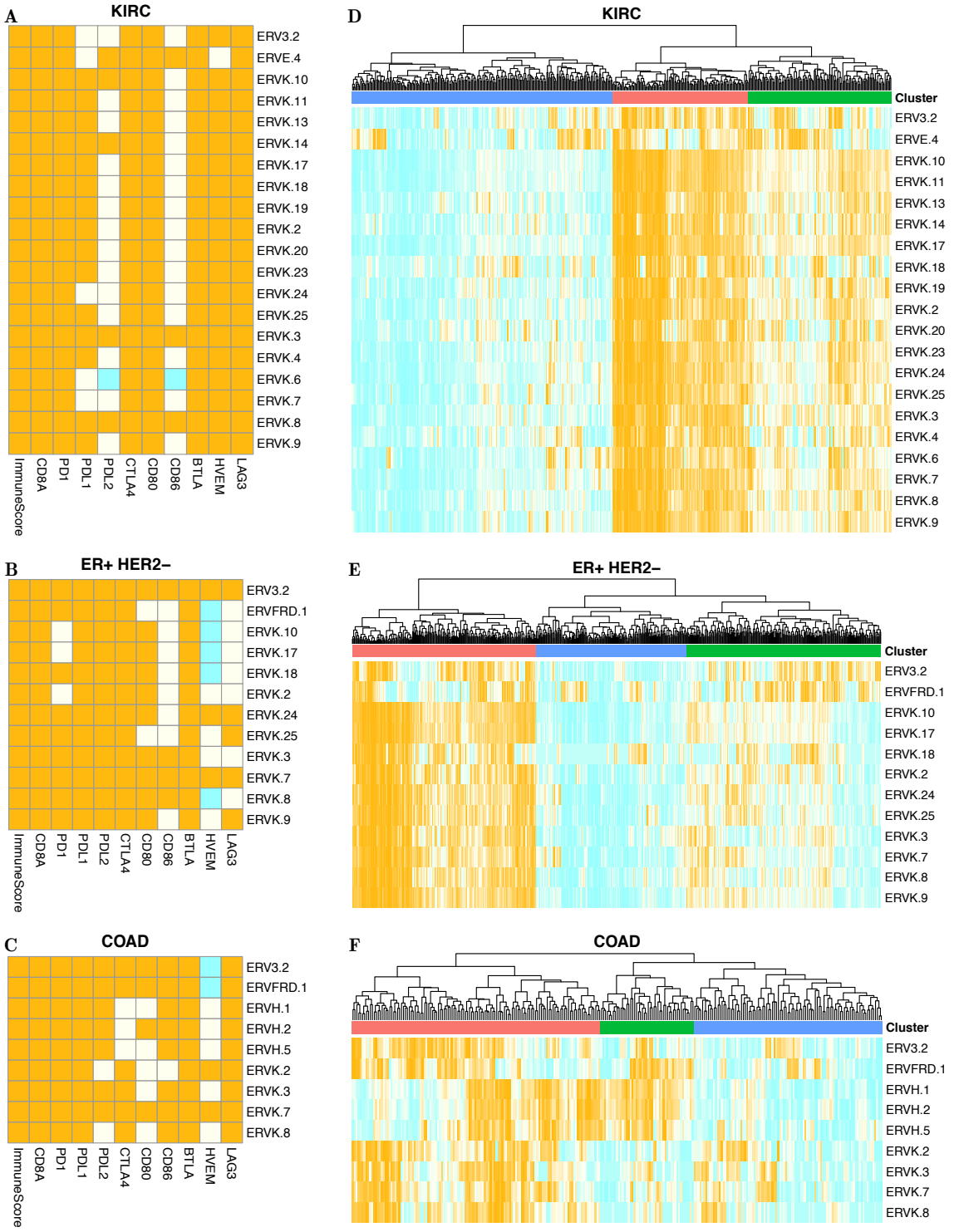


Figure 5.1: Three cancer types with strong evidence of hERV-associated immunogenicity in TCGA.

(A-C) Correlation (Spearman) between expression of potentially immunogenic hERVs (rows) and level of immune markers (columns). Color scheme: discrete, gold: correlated, cyan: anti-correlated, ivory: un-correlated. (D-F) Hierarchical clustering of tumors (columns) by expression (percentile) of potentially immunogenic hERVs (rows). Color scheme: continuous, from cyan (0) to ivory (50) to gold (100). Clusters: red/green/blue = high/intermediate/low group.

5.2.3 Immune activation and checkpoint pathway up-regulation in the high group

In all 3 cancer types, tumors of the high hERV expression (hERV+) group had significantly higher overall immune infiltration into the tumor (panel A of Figure 5.2), higher CD8+ T cell fraction in the infiltrating leukocytes (panel A of Figure 5.2), and higher mRNA expression of CD8A (panel B of Figure 5.2), compared to tumors of the low hERV expression (hERV−) group, showing clear evidence of immune activation. In all 3 cancer types, hERV+ tumors also had significantly higher mRNA expression of checkpoint receptors PD-1 and CTLA-4, and their ligands PD-L1 and CD80 (respectively) compared to hERV− tumors (panel B of Figure 5.2), showing clear evidence of checkpoint (PD-1/CTLA-4) pathway up-regulation. Since hERV+ tumors showed evidence of both immune activation and checkpoint pathway up-regulation, patients with such tumors might respond to immune checkpoint therapy [11].

5.2.4 Immune signatures present in tumor micro-environment of the high group

In addition to CD8+ T cells, M1 macrophages constituted a higher fraction of infiltrating leukocytes in hERV+ tumors compared to hERV− tumors in all 3 cancer types (panel C of Figure 5.2). This relatively increased abundance of M1 macrophage was accompanied by relatively decreased abundance of M0 macrophage in colon cancer and M2 macrophage in clear-cell renal cancer and ER+ HER2− breast cancer (panel C of Figure 5.2). Since M1 macrophages inhibit tumor growth [39] and M2 macrophages promote tumor growth [39], we conclude that hERV+ tumors have a more favorable immune microenvironment. The results also suggest that expression of immunogenic hERVs may induce M1-polarization in tumor infiltrating macrophages. It should be noted that we previously observed such enrichment of M1-polarization in macrophages in case of both hyper-mutation associated immunogenicity and exogenous virus associated immunogenicity.

Additionally, there were some cancer type specific immune signatures: in clear-cell renal cancer, follicular-helper T cells, gamma-delta T cells, activated Natural-Killer cells, resting dendritic cells, and plasma cells constituted a higher fraction of leukocytes in hERV+ tumors compared to hERV− tumors (panel D of Figure 5.2). In colon cancer, activated memory-CD4 T cells constituted a higher

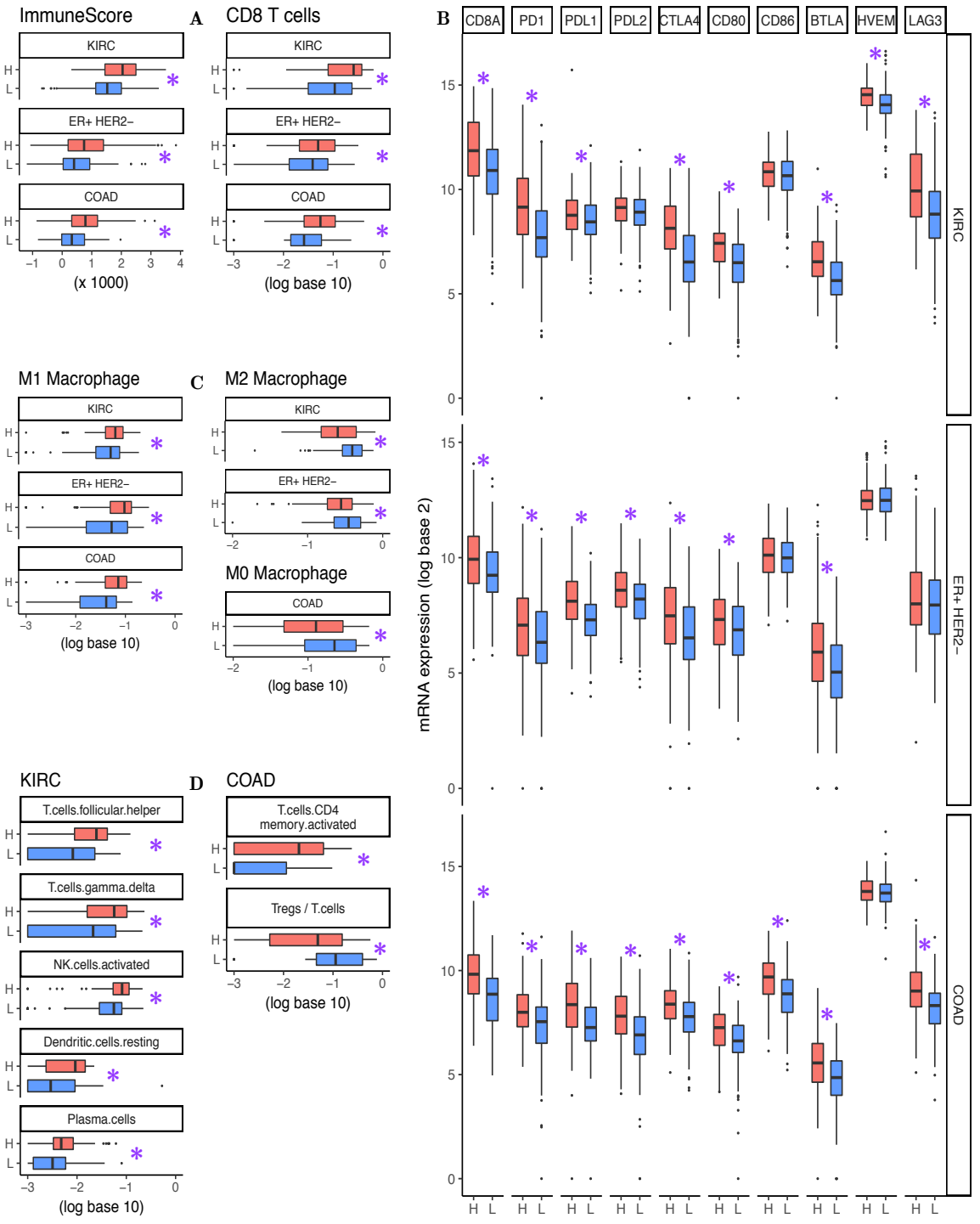


Figure 5.2: Comparison of immune markers between tumors of the high and low groups.

(A) overall immune infiltration in tumor, and CD8+ T-cell fraction in tumor-infiltrating leukocytes. (B) mRNA expression of CD8A and immune checkpoint genes. (C) M1, M2, and M0 macrophages as fraction of leukocytes. (D) Fractional composition of tumor-infiltrating leukocytes, and proportion of T-regs among T cells. * difference is statistically significant (Wilcoxon ranksum test).

fraction of leukocytes, and regulatory T cells constituted a lower fraction of T cells in hERV+ tumors compared to hERV− tumors (panel D of Figure 5.2). Consistent with this observation, anti-correlation between hERV-H expression and intra-tumoral regulatory T-cell density was observed in an independent cohort of colon cancer [75]. Some immune checkpoint genes were also over-expressed in hERV+ tumors compared to hERV− tumors in a cancer type specific manner (panel B of Figure 5.2): examples include BTLA, HVEM, and LAG-3 in clear-cell renal cancer; PD-L2 in ER+ HER2− breast cancer; PD-L2, CD86, and LAG-3 in colon cancer.

5.2.5 Cause of hERV expression in cancers with hERV-associated immunogenicity

To identify the cause of hERV expression in the 3 cancer types, we defined overall hERV expression as the total fraction of RNA-seq reads mapping to 66 transcribed hERVs [71], and searched for genes whose expressions are significantly correlated with overall hERV expression in tumors. 657 genes satisfied this criterion in all 3 cancer types, and enrichment analysis [76] of these genes implicated methyltransferase activity and methylation as the key molecular function and biological process (respectively) associated with overall hERV expression. Figure 5.3 shows the most significantly enriched Gene Ontology (GO) terms from each GO category, where methyl (methyltransferase and methylation) and histone are the most frequent phrases (note the phrases highlighted in blue). These results suggest that epigenetic alteration is the cause of hERV expression in these cancer types. The underlying mechanisms leading to such epigenetic changes remain elusive.

5.2.6 hERV-associated immunogenicity in head-neck squamous-cell cancer

Head-neck squamous-cell cancer (HNSC) showed considerable evidence of hERV-associated immunogenicity (panel A of Figure 5.4). 9 hERVs seemed to be immunogenic, as their expressions were significantly correlated with immune activation (immune infiltration and CD8A expression) and BTLA-HVEM checkpoint pathway expression. Hierarchical clustering of tumors using expression (percentile) of these hERVs stratified the tumors into 2 groups: tumors with high/low expression of potentially immunogenic hERVs (aka hERV+/hERV− tumors). Consequently, hERV+ tumors had

#	ID	Name	FDR
GO: Molecular Function			
1	GO:0018024	histone-lysine N-methyltransferase activity	4.80E-06
2	GO:0042054	histone methyltransferase activity	4.17E-05
3	GO:0001071	nucleic acid binding transcription factor activity	4.17E-05
4	GO:0003700	transcription factor activity, sequence-specific DNA binding	4.17E-05
5	GO:0003723	RNA binding	4.17E-05
6	GO:0016279	protein-lysine N-methyltransferase activity	1.82E-03
7	GO:0016278	lysine N-methyltransferase activity	1.82E-03
8	GO:0042800	histone methyltransferase activity (H3-K4 specific)	1.82E-03
9	GO:0033038	bitter taste receptor activity	5.89E-03
10	GO:0008170	N-methyltransferase activity	5.89E-03
11	GO:0008276	protein methyltransferase activity	5.89E-03
12	GO:0008527	taste receptor activity	1.75E-02
GO: Biological Process			
1	GO:0006397	mRNA processing	5.44E-06
2	GO:0008380	RNA splicing	5.44E-06
3	GO:0018022	peptidyl-lysine methylation	1.78E-05
4	GO:0034968	histone lysine methylation	1.78E-05
5	GO:0016571	histone methylation	1.78E-05
6	GO:0006479	protein methylation	2.93E-05
7	GO:0008213	protein alkylation	2.93E-05
8	GO:0016071	mRNA metabolic process	1.84E-04
9	GO:0031124	mRNA 3'-end processing	1.84E-04
10	GO:0016570	histone modification	2.02E-04
11	GO:0016569	covalent chromatin modification	3.56E-04
12	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	3.56E-04
GO: Cellular Component			
1	GO:0005813	centrosome	2.87E-06
2	GO:0044450	microtubule organizing center part	2.87E-06
3	GO:0005815	microtubule organizing center	3.17E-06
4	GO:0016607	nuclear speck	3.17E-06
5	GO:0005814	centriole	6.19E-06
6	GO:0044451	nucleoplasm part	3.35E-05
7	GO:0035097	histone methyltransferase complex	1.03E-04
8	GO:0015630	microtubule cytoskeleton	9.24E-04
9	GO:0034708	methyltransferase complex	9.24E-04
10	GO:0005681	spliceosomal complex	2.44E-03
11	GO:0097539	ciliary transition fiber	5.90E-03
12	GO:0044441	ciliary part	5.90E-03

Figure 5.3: Enrichment analysis of genes whose expressions are correlated with hERV expression.

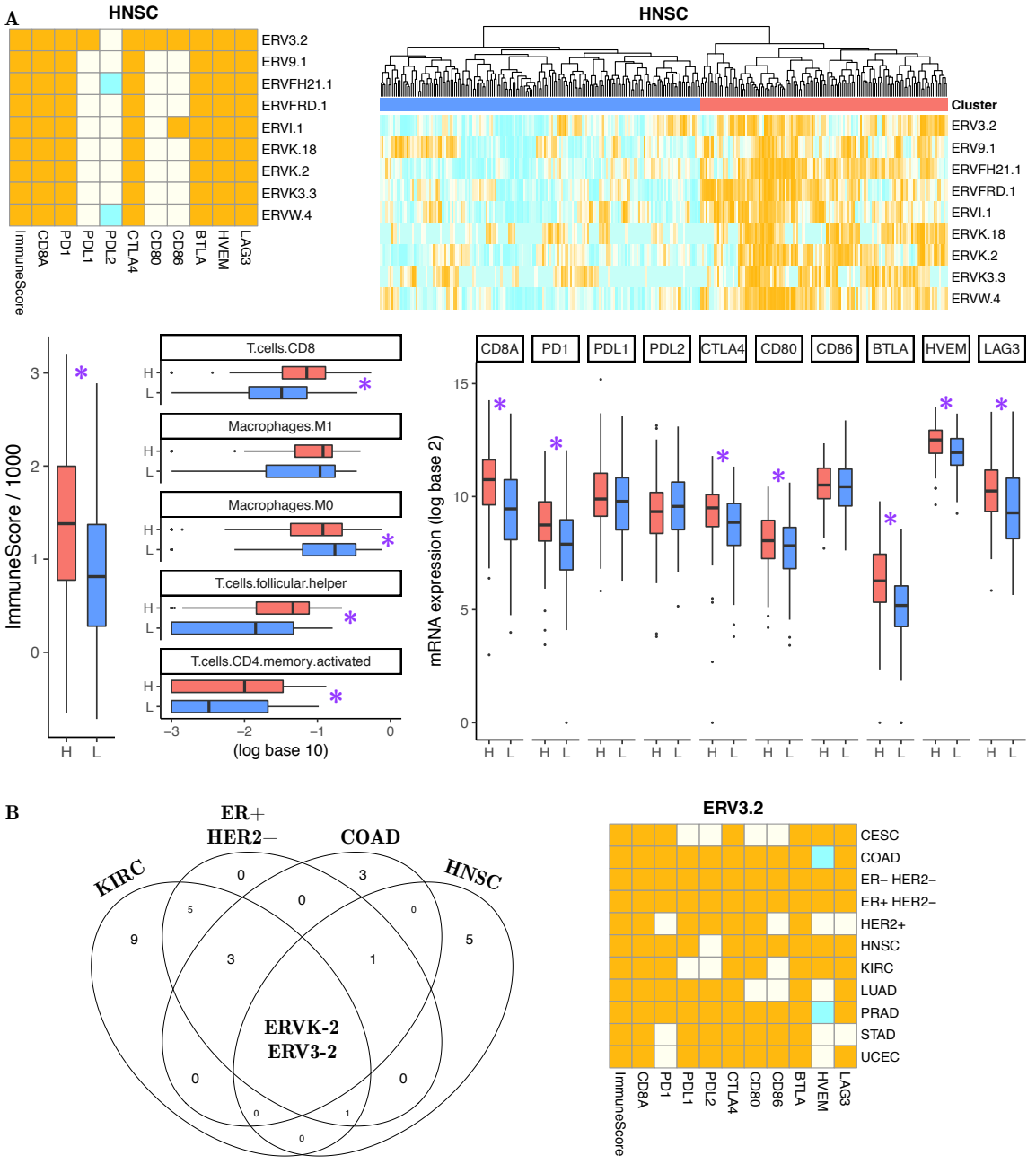


Figure 5.4: Evidence of hERV-associated immunogenicity in other solid cancer types from TCGA.

(A) Evidence in head-neck squamous-cell cancer. Top: identification of potentially immunogenic hERVs, and classification of tumors into high (red) and low (blue) group by expression (percentile) of these hERVs. Bottom: comparison of overall immune infiltration, leukocyte composition, CD8A expression, and expression of immune checkpoint genes between high and low group. (B) hERVs that are potentially immunogenic in multiple cancer types. Left: Venn diagram of potentially immunogenic hERVs, Right: correlation between ERV3-2 expression and level of various immune markers (columns) in 11 solid cancer types (rows). * difference is statistically significant.

significantly higher immune infiltration, CD8+ T cell fraction, CD8A expression, and expression of BTLA and HVEM (and LAG-3) compared to hERV− tumors. Thus, hERV+ HNSC tumors show evidence of both immune activation and checkpoint pathway up-regulation, and may be sensitive to immune checkpoint therapy [11], especially the ones targeting the BTLA-HVEM pathway or LAG-3. However, although PD-1 is over-expressed in hERV+ tumors, neither of its ligands are over-expressed in hERV+ HNSC tumors, making it unclear whether hERV+ HNSC tumors are good candidates for PD-1 blockade.

5.3 Discussion

To summarize, in clear-cell renal cancer, ER+ HER2− breast cancer, colon cancer, and possibly head-neck squamous-cell cancer there is clear evidence of hERV-associated immune activation and checkpoint pathway up-regulation, and hence hERV+ tumors in these cancers may be sensitive to immune checkpoint therapy. 2 hERVs, ERVK-2 and ERV3-2, were found to be immunogenic in all 4 cancer types; and ERV3-2 seemed immunogenic in 7 more solid cancer types (panel B of Figure 5.4). Hence, ERVK-2 and ERV3-2 are the most likely candidates for biomarkers of response to immune checkpoint therapy. Clinical validation is in progress for PD-1 blockade in clear-cell renal cancer.

It should be noted that the results of this analysis do not necessarily suggest the absence of hERV-associated immune checkpoint activation in other cancer types as it was limited by the availability of hERV expression data (only 7 out of 35 cancer types in TCGA had $\geq$250 tumors with both hERV and mRNA expression data). Since the 4 cancer types discussed above all belong to this group of 7 cancer types, there remains the possibility that insufficient sample size prevented the detection of hERV-associated immunogenicity in some of the remaining cancer types.

Chapter 6

Future directions

We have seen that several mechanisms can lead to immune recognition of a tumor resulting in an anti-tumor immune response. If the tumor blocks this response using immune checkpoint pathways, it makes the tumor vulnerable to immune checkpoint therapy. We have discussed two general mechanisms of response to immune checkpoint therapy in the last 4 chapters: (1) hyper-mutation in tumor due to exogenous mutagenic exposure, or intrinsic defect in proofreading or mismatch repair of DNA, or aberrant APOBEC activity; (2) expression of exogenous or endogenous viral RNA in tumor. Using new data from various public data sources and clinical data from patients treated with immune checkpoint therapy at R-CINJ and VICC, the next step would be to extend these studies to address several questions that remain open, as explained in the next 3 sections.

6.1 Questions related to hyper-mutation

6.1.1 Is a defect in homologous recombination by itself a potential biomarker?

Defects in homologous recombination (HR), found in 31% ovarian cancer, 27% breast cancer, 7.5% pancreatic cancer, and 7% gastric cancer [77], can lead to hyper-mutation in tumors. But while the etiologies mentioned above cause an abundance of single base substitutions and small indels, HR defects cause an abundance of large indels and structural rearrangements in tumors, with few small indels and single base substitutions. In TCGA whole exome sequencing dataset and various whole genome sequencing datasets, it is interesting to test whether this type of hyper-mutation is also associated with immune checkpoint activation, and may be a potential biomarker of response to immune checkpoint therapy.

6.1.2 What are the driver genes in hyper-mutant tumors of above etiologies?

Given the large number of genes that are mutated in hyper-mutant tumors, computational algorithms often find it difficult to identify driver alterations in hyper-mutant tumors. For example, analysis using the commonly used MutSigCV [44] algorithm was unable to identify driver mutations in POLE mutant endometrial cancer. Using high depth clinical sequencing data recently made available by Memorial Sloan Kettering Cancer Center and Foundation Medicine, it should be possible to identify and validate both tissue type specific and tissue type independent driver genes in hyper-mutant tumors of each etiology mentioned above.

6.2 Questions related to viral mechanisms

6.2.1 Is over-expression of LINE/SINE elements associated with response to immune checkpoint therapy?

As a follow-up of immune checkpoint activation from hERVs, which are Long Terminal Repeats (LTR) retrotransposons, it will be of interest to perform a pan-cancer analysis of the TCGA dataset to test whether expression of any non-LTR retrotransposons in tumor is associated with immune checkpoint activation in any cancer type. Non-LTR retrotransposons include long interspersed elements (LINEs) and short interspersed elements (SINEs), that constitute 21% and 13% of human genome respectively [69], but are not expressed in normal somatic tissues (for comparison, protein coding regions constitute only 1% of the human genome [70]). Using clinical and experimental data from collaborators at R-CINJ and VICC, it should also be possible to validate expression of the identified LINE/SINE element as a biomarker of response to immune checkpoint therapy in the identified cancer type.

6.2.2 Is hypo-methylation a relevant functional mechanism in this context?

Hypo-methylation is a potentially functional mechanism responsible for retrotransposon expression in tumors. Consequently, a pan-cancer analysis of the TCGA dataset to identify the cancer types

where hypo-methylation is associated with immune checkpoint activation is of interest. If evidence of hypo-methylation-associated immunogenicity is found in a cancer type where clinical data from patients treated with immune checkpoint therapy is available, methylation burden can be tested as a biomarker of response to immune checkpoint therapy in that cancer type. In addition, the association between hypo-methylation and over-expression of hERVs, LINEs, SINEs, and lincRNAs can be tested to see if these two are cause and effect.

6.3 Questions related to the basic hypothesis

6.3.1 Is it possible to develop a small assay for early prediction of eventual response?

Exploiting the concept that CD8+ T cell infiltration in tumor induces upregulation of immune checkpoint pathways, it may be possible to construct a small assay (15 genes) and apply it to a recently published dataset [78] of patients undergoing immune checkpoint therapy. The goal would be to test whether, during the initiation phase of immune checkpoint therapy (after 1-2 month of treatment), such an assay can reliably distinguish patients who will eventually respond to continued treatment from those who will not.

6.3.2 Investigate mechanisms of constitutive PD-L1 expression or constitutive loss of PD-L1 expression, and their clinical implication on response or resistance

In addition to CD8+ T cell infiltration in tumor, several other causes of PD-L1 expression have recently been proposed [79]. A pan-cancer analysis of the TCGA dataset would determine which of these proposed genetic processes (if any) are associated with constitutive PD-L1 expression. Using data from 2,000 tumors sequenced in the Precision Medicine Initiative (PMI) at R-CINJ, it should be possible to identify tumors that display those genetic characteristics, followed by clinical testing by our pathology colleagues, who can quantify PD-L1 expression and lymphocyte infiltration in those tumors. If some of these patients received PD-1 blockade, it will also be possible to test therapeutic implication of constitutive PD-L1 expression or constitutive loss of PD-L1 expression on response or resistance to PD-1 blockade.

Bibliography

[1] J. M. Mehnert, *et al.*, *J. Clin. Invest.* **126**, 2334 (2016).

[2] A. Panda, *et al.*, *JCO Precision Oncology* pp. 1–13 (2017).

[3] A. Panda, *et al.*, *J. Natl. Cancer Inst.* (2017).

[4] C. L. Chaffer, R. A. Weinberg, *Science* **331**, 1559 (2011).

[5] P. Mehlen, A. Puisieux, *Nat. Rev. Cancer* **6**, 449 (2006).

[6] J. W. Kim, J. P. Eder, *Oncology (Williston Park, N.Y.)* **28 Suppl 3**, 15 (2014).

[7] Ö. Türeci, *et al.*, *Clinical Cancer Research* **22**, 1885 (2016).

[8] V. K. Anagnostou, J. R. Brahmer, *Clinical Cancer Research* **21**, 976 (2015).

[9] A. Ribas, *N. Engl. J. Med.* **373**, 1490 (2015).

[10] D. M. Pardoll, *Nat. Rev. Cancer* **12**, 252 (2012).

[11] P. C. Tumeh, *et al.*, *Nature* **515**, 568 (2014).

[12] A. Snyder, *et al.*, *N. Engl. J. Med.* **371**, 2189 (2014).

[13] E. M. Van Allen, *et al.*, *Science* **350**, 207 (2015).

[14] D. B. Johnson, *et al.*, *Cancer Immunol Res* (2016).

[15] N. A. Rizvi, *et al.*, *Science* **348**, 124 (2015).

[16] D. T. Le, *et al.*, *N. Engl. J. Med.* **372**, 2509 (2015).

[17] J. E. Rosenberg, *et al.*, *Lancet* **387**, 1909 (2016).

[18] P. T. Nghiem, *et al.*, *N. Engl. J. Med.* **374**, 2542 (2016).

[19] H. L. Kaufman, *et al.*, *Lancet Oncol.* **17**, 1374 (2016).

[20] Y.-L. Kwong, *et al.*, *Blood* **129**, 2437 (2017).

[21] L. B. Alexandrov, *et al.*, *Nature* **500**, 415 (2013).

[22] L. B. Alexandrov, *Science* **350**, 1175 (2015).

[23] L. B. Alexandrov, *et al.*, *Nat. Genet.* **47**, 1402 (2015).

[24] M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, *Cell* **160**, 48 (2015).

[25] K. Yoshihara, *et al.*, *Nat Commun* **4**, 2612 (2013).

[26] A. M. Newman, *et al.*, *Nat. Methods* **12**, 453 (2015).

[27] A. H. Ramos, *et al.*, *Hum. Mutat.* **36**, E2423 (2015).

[28] Cancer Genome Atlas Network, *Nature* **490**, 61 (2012).

[29] Cancer Genome Atlas Research Network, *Nature* **513**, 202 (2014).

[30] E. Cerami, *et al.*, *Cancer Discov* **2**, 401 (2012).

[31] J. Gao, *et al.*, *Sci Signal* **6**, pl1 (2013).

[32] Cancer Genome Atlas Research Network, *et al.*, *Nature* **497**, 67 (2013).

[33] E. Heitzer, I. Tomlinson, *Curr. Opin. Genet. Dev.* **24**, 107 (2014).

[34] Y. R. Hussein, *et al.*, *Mod. Pathol.* **28**, 505 (2015).

[35] M. Khunger, *et al.*, *JCO Precision Oncology* pp. 1–15 (2017).

[36] I. C. van Gool, *et al.*, *Clinical Cancer Research* **21**, 3347 (2015).

[37] S. Bellone, *et al.*, *Gynecol. Oncol.* **138**, 11 (2015).

[38] I. Waldhauer, A. Steinle, *Oncogene* **27**, 5932 (2008).

[39] A. Yuan, *et al.*, *Sci Rep* **5**, 14273 (2015).

[40] B. E. Howitt, *et al.*, *JAMA Oncol* **1**, 1319 (2015).

[41] J. Gong, C. Wang, P. P. Lee, P. Chu, M. Fakih, *J Natl Compr Canc Netw* **15**, 142 (2017).

[42] T. M. Johanns, *et al.*, *Cancer Discov* **6**, 1230 (2016).

[43] E. Bouffet, *et al.*, *J. Clin. Oncol.* **34**, 2206 (2016).

[44] M. S. Lawrence, *et al.*, *Nature* **499**, 214 (2013).

[45] J. T. Robinson, *et al.*, *Nat. Biotechnol.* **29**, 24 (2011).

[46] H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, *Brief. Bioinformatics* **14**, 178 (2013).

[47] E. R. Fearon, B. Vogelstein, *Cell* **61**, 759 (1990).

[48] B.-K. Koo, *et al.*, *Nature* **488**, 665 (2012).

[49] E. Deacu, *et al.*, *Cancer Research* **64**, 7690 (2004).

[50] S. Rao, J. E. Stadanlick, K. Q. Cai, D. L. Wiest, *Cancer Research* **75**, 5215 (2015).

[51] D. E. Brash, *Photochem. Photobiol.* **48**, 59 (1988).

[52] T. M. Hernandez-Boussard, P. Hainaut, *Environ. Health Perspect.* **106**, 385 (1998).

[53] D. T. Le, *et al.*, *Science* **357**, 409 (2017).

[54] S. A. Roberts, *et al.*, *Nat. Genet.* **45**, 970 (2013).

[55] T. Sorlie, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8418 (2003).

[56] W. Zou, *Nat. Rev. Immunol.* **6**, 295 (2006).

[57] R. Nanda, *et al.*, *J. Clin. Oncol.* **34**, 2460 (2016).

[58] R. J. Motzer, *et al.*, *N. Engl. J. Med.* **373**, 1803 (2015).

[59] S. A. Shukla, *et al.*, *Nat. Biotechnol.* **33**, 1152 (2015).

[60] D. Marvel, D. I. Gabrilovich, *J. Clin. Invest.* **125**, 3356 (2015).

[61] K. B. Chiappinelli, *et al.*, *Cell* **162**, 974 (2015).

[62] D. Roulois, *et al.*, *Cell* **162**, 961 (2015).

[63] K. Muro, *et al.*, *Lancet Oncol.* **17**, 717 (2016).

[64] K. L. Grogg, *et al.*, *Mod. Pathol.* **16**, 641 (2003).

[65] A. M. Chiaravalli, *et al.*, *Virchows Arch.* **448**, 344 (2006).

[66] S. Derks, *et al.*, *Oncotarget* **7**, 32925 (2016).

[67] C. Ma, *et al.*, *Am. J. Surg. Pathol.* p. 1 (2016).

[68] A. Kawazoe, *et al.*, *Gastric Cancer* pp. 1–9 (2016).

[69] E. S. Lander, *et al.*, *Nature* **409**, 860 (2001).

[70] S. B. Ng, *et al.*, *Nature* **461**, 272 (2009).

[71] J. Mayer, J. Blomberg, R. L. Seal, *Mob DNA* **2**, 7 (2011).

[72] O. Hohn, K. Hanke, N. Bannert, *Frontiers in Oncology* **3**, 246 (2013).

[73] R. Löwer, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4480 (1993).

[74] H. L. Wu, *et al.*, *Retrovirology* **13**, 6 (2016).

[75] N. Desai, *et al.*, *JCI Insight* **2**, e91078 (2017).

[76] J. Chen, E. E. Bardes, B. J. Aronow, A. G. Jegga, *Nucleic Acids Res.* **37**, W305 (2009).

[77] L. B. Alexandrov, *et al.*, *Nat Commun* **6**, 8683 (2015).

[78] P.-L. Chen, *et al.*, *Cancer Discov* **6**, 827 (2016).

[79] A. Ribas, S. Hu-Lieskovan, *J. Exp. Med.* **213**, 2835 (2016).

List of Abbreviations

ACC : Adrenocortical Carcinoma

BLCA : Bladder Urothelial Carcinoma

ccRCC : clear-cell Renal Cell Carcinoma

CESC : Cervical Squamous Cell Carcinoma

CHOL : Cholangiocarcinoma

COAD : Colon Adenocarcinoma

EBV : Epstein Barr Virus

ER+ HER2− : a subtype of Breast Cancer

ER− HER2− : a subtype of Breast Cancer

ESCA : Esophageal Carcinoma

FFPE : Formalin Fixed Paraffin Embedded

GBM : Glioblastoma Multiforme

H&E : Hematoxylin and Eosin

HBV : Hepatitis B Virus

HCV : Hepatitis C Virus

HER2+ : a subtype of Breast Cancer

hERV : human Endogenous Retrovirus

HNSC : Head-Neck Squamous-cell Carcinoma

HPV : Human Papillomavirus

iCAM : immune Checkpoint Activating Mutation burden

KICH : Kidney Chromophobe

KIRC : Kidney Renal Clear Cell Carcinoma

KIRP : Kidney Renal Papillary Cell Carcinoma

LGG : Brain Lower Grade Glioma

LIHC : Liver Hepatocellular Carcinoma

LUAD : Lung Adenocarcinoma

LUSC : Lung Squamous Cell Carcinoma

MESO : Mesothelioma

MSI : Microsatellite Instable

MSS : Microsatellite Stable

ORF : Open Reading Frame

OV : Ovarian Serous Cystadenocarcinoma

PAAD : Pancreatic Adenocarcinoma

PCPG : Pheochromocytoma and Paraganglioma

PRAD : Prostate Adenocarcinoma

R-CINJ : Rutgers Cancer Institute of New Jersey

READ : Rectum Adenocarcinoma

SARC : Sarcoma

SKCM : Skin Cutaneous Melanoma

STAD : Stomach Adenocarcinoma

TCGA : The Cancer Genome Atlas

TGCT : Testicular Germ Cell Tumors

THCA : Thyroid Carcinoma

THYM : Thymoma

UCEC : Uterine Corpus Endometrial Carcinoma

UCS : Uterine Carcinosarcoma

UVM : Uveal Melanoma

VICC : Vanderbilt Ingram Cancer Center