# BIG DATA ANALYTICS AND CANCER BIOLOGY: LESSONS IN TAKING SCIENCE FROM THE PROCESSOR TO THE PATIENT.

by

Nicholas Craig Borcherding

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Biomedical Sciences (Cancer Biology) in the
Graduate College of
The University of Iowa

May 2020

Thesis Supervisor: Associate Professor Weizhou Zhang

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

PH.D. THESIS

---

This is to certify that the Ph.D. thesis of

Nicholas Craig Borcherding

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Biomedical Sciences (Cancer Biology) at the May 2020 graduation.

Thesis Committee: _____
Weizhou Zhang, Thesis Supervisor

_____
Adam Dupuy

_____
Munir Tana

_____
Ronald Weigel

_____
Frank Zhan

To my wife, my best friend, the only person I would ever trust on a time travel mission.

God will not have his work made manifest by cowards. A man is relieved and gay when he has put his heart into his work and done his best; but what he has said or done otherwise, shall give him no peace. It is a deliverance which does not deliver. In the attempt his genius deserts him; no muse befriends; no invention, no hope.

Ralph Waldo Emerson, Self-Reliance

**ACKNOWLEDGEMENTS**

There are many people and things that I should acknowledge First, my wife Kirsten, for putting up with my eccentricities. I will be endlessly thankful to my research mentors: Matt, Kevin, Weizhou and Ryan for helping me build a reality in which I am truly happy. That same thankfulness extends to Pam, for helping me in the moment in which I lost my true happiness. There have been many friends, peers, and professors that have made an impact on me while in medical and graduate school and I wouldn't be here without them. Ultimately, I must acknowledge my mother and father, it is their fault that I am me, genetically and epigenetically. After pouring my heart into writing my Master's thesis, I gave them both bound copies, for which they told me it they were proud of me and couldn't understand any of it. I didn't expect their lack of understanding to extend that far beyond my teenage years and the experience largely informed my approach to scientific communication.

## ABSTRACT

The immune system provides a harsh selection pressure for tumors to develop mechanisms to evade destruction. Tumors can function to subvert adaptive immune activation by decreasing expression of major histocompatibility complex I molecules to prevent the activating signal or through directing the T cell differentiation and activity. The latter is a diverse set of methods involving the interaction of cells in the tumor microenvironment and can function by increasing infiltration of immunosuppressive regulatory T cells (Tregs), myeloid-derived suppressor cells, decreasing antigen presentation, the secretion of suppressive cytokines/chemokines, or the overexpression of negative regulators of immune response, so-called immune checkpoints. Therapies which elicit an anti-tumor immune response, either by directly stimulating an immunogenic response or targeting inhibitory pathways, have long been sought after. My studies focus on promoting or directing the immune response by overcoming the tumor-mediated suppression and increasing response to immunotherapies.

A major predictive correlate of immune checkpoint blockades is tumor mutational load. Although possessing high levels of genomic instability, the aggressive basal-like breast cancer does not respond well to immune checkpoint blockade. My work identifies both an increased level and poor prognostic indication of the DNA mismatch repair proteins, MSH2 and MSH6, in basal-like breast cancer. Lynch syndrome is a familial disorder with germline defects in DNA mismatch repair, a hallmark of these tumors is dense immune infiltration. Conversely, we found the increasing level of MSH2 protein in basal-like breast cancer was associated with decreased immune signatures, notably in lymphocyte, NK cell, and myeloid cell signatures. The genetic ablation of *Msh2* in basal-like breast cancer models led to significant reduction in tumor growth and an increase in survival time. The addition of the anti-PD-1 immunotherapy led to a further reduction in tumor growth and increase in T-cell infiltration in tumors from *Msh2* knockout *versus Msh2* wild-type cells. This work indicates altering the DNA repair process may act as an adjuvant for tumor immunotherapy and increase efficacy of immune checkpoint blockade in aggressive breast cancers.

Another critical component of the tumor microenvironment is the suppressive pressure of Tregs. Analysis of single-cell RNA sequencing of immune cells from both renal and hepatocellular tumors found heterogeneity among Tregs, with tumor-infiltrating Tregs containing distinct, overlapping expression patterns between the two cancer types. Further analysis of other datasets identified several genes commonly differentially regulated in tumor-infiltrating Tregs compared to peripheral-blood Tregs, one of which was *CD177*. We confirmed the presence of CD177 on of tumor-infiltrating Tregs, but not peripheral-blood Tregs, in humans and found increased expression of several suppressive markers and chemokine receptors on $CD177^+$ breast and renal cancer-infiltrating Tregs compared to $CD177^-$ Tregs. Initial human and mouse studies found that tumor-

infiltrating CD177$^{+}$ Tregs were more suppressive than CD177$^{-}$ Tregs in *ex vivo* suppression assays. Furthermore, we found the removal of *Cd177* led to decrease tumor growth in both breast and colorectal mouse models. Taken together, this work lays a foundation for better targeting of tumor-infiltrating Tregs by identifying novel markers of suppressive Treg subsets.

Beyond characterization of immune cell infiltrates of solid tumors, my work also examines the heterogeneity within blood malignancies. Cutaneous T cell lymphomas (CTCL), encompassing a spectrum of T-cell lymphoproliferative disorders involving the skin, have collectively increased in incidence over the last 40 years. Sézary syndrome (SS) is an aggressive form of CTCL characterized by significant presence of malignant cells in both the blood and skin. The guarded prognosis for SS reflects a lack of reliably effective therapy, due in part to an incomplete understanding of disease pathogenesis. Using single-cell sequencing of RNA, we confirmed that SS is a clonal disease by virtue of shared T-cell receptor VDJ expression and CDR3 sequence, but we further defined a more complex model featuring distinct transcriptomic states within SS. Furthermore, we developed methodologies to utilize the transcriptomic diversities in SS to predict disease stage. This work offers insight into the heterogeneity of SS, providing better understanding of the transcriptomic diversities within a clonal tumor, which can predict tumor stage and thereby offer guidance of therapy.

Although a diverse set of projects, my studies focus on the axis of immune-tumor interaction and the development of computational methods to identify targets to improve immunotherapies. In finding the poor prognostic indication of DNA mismatch repair constituents in basal-like breast cancer, my work identifies MSH2 and MSH6 as novel tumor promoters in the context of immune evasion. Furthermore, the analysis of transcriptional heterogeneity of tumor-infiltrating Tregs is the first demonstration of unique gene expression patterns of Tregs in tumors, which can be used to predict survival. Similarly, my work shows the heterogeneity of CTCL cells could be a major underlying cause of the development resistance and poor outcomes in the treatment of advanced SS.

## PUBLIC ABSTRACT

If I may borrow terminology from my time in the Marine Corps infantry, cancer is a war on oneself. For the last 50 years, oncologists have fought this war with broad, non-specific strategies like chemotherapy and radiation that cause collateral damage to the individual patient and lasting repercussions on health. More recently, the dawn of targeted therapies is like introducing snipers to the battlefield, increasing the accuracy of the therapy against tumors and reducing the side effect profiles of cancer treatments. However, targeted therapies generally have a limited effective window of time, as the tumor develop tactics to avoid the therapies, what is termed resistance. In the modern era of cancer combat, an emerging strategy is using the immune system to fight the tumor. My work focuses on how to arm and direct the immune system against a tumor, limiting the collateral damage of chemotherapies and resistance of targeted therapies. In this sense, having the immune system fight cancer functions to produce a force that can evolve with the tactics that the cancer might develop, leading to more profound, long-term responses. The limitation currently in this strategy is identifying patients that may benefit from the immunotherapy and trying to better activate the immune system for patients that would not respond initially.

**TABLE OF CONTENTS**

## LIST OF TABLES

**Table**

## LIST OF FIGURES

**Figure**

# CHAPTER 1
## INTRODUCTION

About a year into my graduate studies, I was at a reception to celebrate the completion of the Liaison Committee for Medical Education re-accreditation process, a large, bureaucratic process that occurs every seven years for medical schools. While having dinner, the conversation turned to discussion of probability. One of the doctors stated he was less of a clinician than a Bayesian statistician, dealing with probabilities and patients. Bayesian statistics is a subfield of statistics derived from the work of the minister Thomas Bayes who attempted to predict the likelihood of winning a lottery based on prior knowledge. In other words, Bayesian statistics deals in epistemic confidence, estimating the likelihoods of reality. On this vein of thinking, diagnostic heuristics that this doctor uses are based on the evidence the patient presents and the prior knowledge of the likelihood of the competing diagnoses. This idea is referred to as Occam's Razor, which is based on the work of William of Ockham, a Franciscan friar, and states among competing hypotheses, the one with the fewest assumptions should be selected. A common shorthand for this concept in clinic is, 'look for horses, not zebras' to express the most likely solution is the probable solution. The use of diagnostic parsimony was expanded throughout the works of Sir Arthur Conan Doyle, a doctor himself, who modeled his infamous Sherlock Holmes on one of his medical school professors. In the second novel of the series, The Sign of Four, Sherlock explains to his friend Dr. James Watson, "when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth?"[1] Medicine , research, and Detective Holmes rely on objective realities and assumptions, the latter being prone to our cognitive biases.

In the same year as my dinner, Richard Thaler, an economist from the University of Chicago, won the Nobel Memorial Prize in Economic Science for his work in the founding of behavioral economics. Thaler, along with Amos Tversky and Danny Kahnamen, introduced psychology to economics, characterizing the lack of rationality in decision making. In 1972, the iconoclastic Tversky and Kahneman, demonstrated how the *ad hoc* Bayesian Statistician in all of us is inherently faulty. Termed the representativeness heuristic, or how the degree of similarity weighs probability estimates for the worse. Through the course of the next 20 years, Thaler, Kahneman, and Tversky would demonstrate the irrationality of human decision making. The apostatic financial analyst turned philosopher, Nassim Taleb, would later apply a more nihilistic conclusion to human decision making in his 2007 book, The Black Swan. In the book, Taleb discusses unexpected events have orders of magnitude greater influence than compared highly-reproducible predictions of expected events.[2]

How does economic theory apply to cancer biology? The simplistic answer focuses on the need for better statistical approaches and understanding of limitations as the wave of highthroughput technologies become a mainstay in clinical medicine and research. But like many scientific fields, there is much more nuisance. There

is a necessary balance that needs to be achieved in using big data analytics. A balance between using large data sets as a tool to reveal unique biological understanding and the tendency for overfitting and overinterpretation of results. This phenomenon of conclusions based on extrapolation of small numbers or nonrepresentative sampling is the basis of erroneous or misleading results that are found in classic phase I or phase II clinical trials. The tendency to overfit data is not unique in medical disciplines and is partly responsible for the current issues with reproducibility in research. One key to balance the demands of large data analysis is radical transparency, making both raw data and code available for other researchers. Additionally, having researchers and trainees receive education in data science is a requisite step for the new age of cancer computational biology.

For my part, this dissertation will be a diverse study in how to use computational approaches in cancer research with a weighted interest in cancer immunology. Each chapter examines a unique aspect of cancer biology with a particular interest in using big data to reveal new insights in both translational and biological phenomena. Hopefully, by beginning this dissertation by discussing the faulty statistical inferences that human brains make, I hope I have properly framed the reader to expect more pronounced discussion on statistical method and limitations of the computational approaches. Fittingly, Amos Tversky and Danny Kahneman first characterized the framing effect of human decision making in one of their more famous Science articles in 1981.[3]

**Breast cancer and big data, a case study.**

Cancer is singularly ominous as a term. Not only because the word invokes the feelings of family members and friends that have suffered, but because it represents the betrayal of one's self. Among cancers, breast cancer is held in the collective consciousness for a number of reasons. Most prominently to this collective concern for breast cancer is the overall prevalence. In the United States, nearly 281,000 new diagnoses of breast cancer will be made this year and 43,000 patients will succumb to their disease.[4] With the current incidence rates, 1 in 8 women in the United States will be diagnosed with breast cancer in their lifetime, making it, to borrow Mr. Holmes's words, *improbable* that breast cancer will not in some way affect you or a loved one.

Breast cancer is a spectrum of diseases that, like other cancers, is named for the tissue of origin rather than the underlying clinical and molecular characteristics. In the interest of demonstrating the use of diagnostic heuristics, we will start to demarcate the subclassifications of breast cancer using a similar methodology as a clinician. These subclassifications are based on histological type and biomarker status or gene expression patterns, which are used to identify therapies and predict the course of the disease. The current screening guidelines for breast cancer from the United States Preventative Task Force are biennial mammography for women ages 50-74.[5] While screening before the age of 50, clinicians and patients should take into account family history and a patient's values. A special exception is made for familial cancer syndromes, like *BRCA1/BRCA2* mutations, or previous history of radiation; these patients should receive annual screenings

starting at 30-40 years of age, depending on their mutation.[6] Abnormal findings occur in roughly 10% of women during the screening mammography.[7,8] In these cases, a secondary diagnostic bilateral mammography is recommended.[6] As a caveat, of these women getting secondary mammography, only 5% will be diagnosed with cancer, which translates to a 50% risk of a false positive over the course of 10 years for every woman.[7,8]

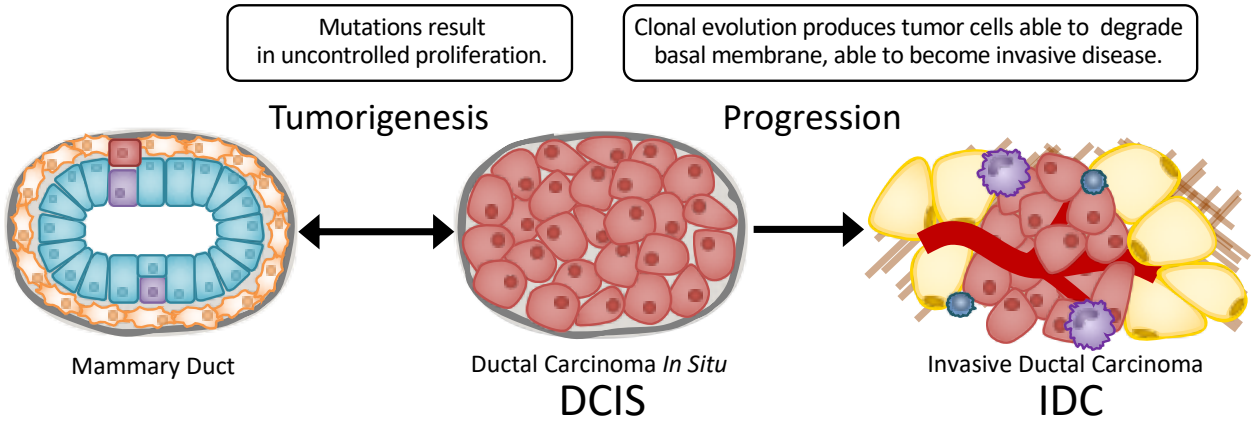


**Figure 1.** Invasive ductal carcinoma tumor initiation and progression. Oncogenic hits lead to a proliferation of mammary ductal cells leading to ductal carcinoma *in situ*. As non-obligate precursors lesions can regress or progress to invasive disease.

After a positive diagnostic mammography, there are several steps in determining the type and sub-classification of breast cancer. Mammography, in addition to other imaging modalities, like ultrasound or magnetic resonance imaging, can be used to determine if the breast cancer is invasive, a cancer that has spread beyond the layers of the tissue of origin, or a carcinoma *in situ* (CIS), referring to cancer cells that reside within the defined structures of the tissue. As the description suggest, CIS is considered a non-obligate precursor lesion to invasive breast cancer, with an estimated 20-50% of CIS progressing to invasive disease (**Figure 1**).[9] In the context of nearly 281,000 diagnosis of invasive breast cancer per year, CIS is diagnosed in over 60,000 women per year in the United States.[4,10] The next and vital step in the diagnostic framework for breast cancer is the biopsy and histopathological examination, which may result from a fine-needle aspiration or from resection of the tumor itself.[6] Invasive breast cancer and CIS is principally comprised of either ductal (approximately 80% of diagnoses), referring to the resemblance of the tumor cells of the mammary duct, or lobular carcinoma (approximately 10-15%), referring to the resemblance of the milk-producing structures of the mammary gland.[10] Other rarer types of invasive breast cancer that can be revealed upon histological examination include: medullary (2-5%), mucinous (1-6%), tubular (1-4%), inflammatory (1-2%), and papillary (0.5%).[11–15] These incidence rates form the backbone of the diagnostic heuristic for breast cancer with invasive ductal carcinoma (IDC) being the most likely diagnosis, followed by invasive lobular carcinoma (ILC) or ductal carcinoma *in situ* (DCIS).

Clinically, IDC can be further differentiated by the presence or absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) via immunohistochemical (IHC) staining. Due to the high level of copy number amplification of the *ERBB2* gene locus for HER2, chromogenic or fluorescent *in situ* hybridization (CISH/FISH) is also often employed. The staining patterns generate three principal subtypes, 1) $ER^+$, 2) $HER2^+$, and 3) triple-negative breast cancer (TNBC), lacking ER, PR, and HER2. These clinical designations predict the course of disease with TNBC and $HER2^+$ tumors possessing worse prognoses with increased risk of recurrence and metastasis compared to $ER^+$ breast cancers.[16–19] Additionally, the presence of ER and HER directs the use of Food and Drug Administration (FDA)-approved targeted therapies, like selective estrogen receptor modulators (SERMs), cyclin-dependent kinase (CDK) inhibitors for $ER^+$ tumors, and monoclonal antibodies or small molecule inhibitors that target HER2. Similar to the IHC staining, the emergence of highthroughput technologies have begun to change the clinical management of patients diagnosed with breast cancer and the understanding of the molecular/genetic underpinnings that drive specific cancers.

In a series of papers, Charles Perou and Therese Sørlie demonstrated the importance of gene expression patterns in understanding and stratifying breast cancer.[20,21] Using microarray quantifications of breast tumor transcriptomes, they identified molecular subtypes based on hierarchical clustering. Initial classifications from their work included basal-like breast cancer (BLBC), *ERBB2*$^+$($HER2^+$), luminal A, luminal B, and normal breast-like (**Figure 2**). Using the clinical subtypes as a foil, these five molecular subtypes essentially represent stratifications of the clinical subtypes. TNBC can be thought of as either basal-like (70-80% of the cases) or normal-like (20-30% of cases).[22] Basal-like refers to the expression of basal-cell-associated cytokeratins (KRT5/6/14/17) and epidermal growth factor receptor 1 (EGFR) and account for the poorest prognosis in the TNBC patient population.[22,23] Likewise, $ER^+$ tumors can be divided into Luminal A or Luminal B, referring to their staining of luminal cytokeratins 8 and 18. Luminal A breast cancers are generally $ER^+PR^+HER2^-$ and are generally responsive to SERMs or aromatase inhibitors, while Luminal B breast cancers have variable expression of ER, PR, and HER2 with a more aggressive disease course.[20] Using this molecular subtyping work as a basis, a follow-up study utilized a centroid-based machine learning prediction method, named Prediction Analysis of Microarray (PAM), to identify 50 genes that could differentiate BLBC, $HER2^+$, Luminal A, and Luminal B breast cancers.[24] The PAM approach utilizes centroid, or average gene expression in a given class, to produce a t-statistic weighted by the mean difference between the groups. Ultimately, PAM identifies stable and significantly-differential genes with large differences in mean expression.[24,25] These 50 genes that differentiate the four major molecular subtypes were later developed into an polymerase-chain reaction-(PCR-) based assay, called the Prosigna Breast Cancer Prognostic Gene Signature Assay or the PAM50 and received FDA clearance in 2013.[26]

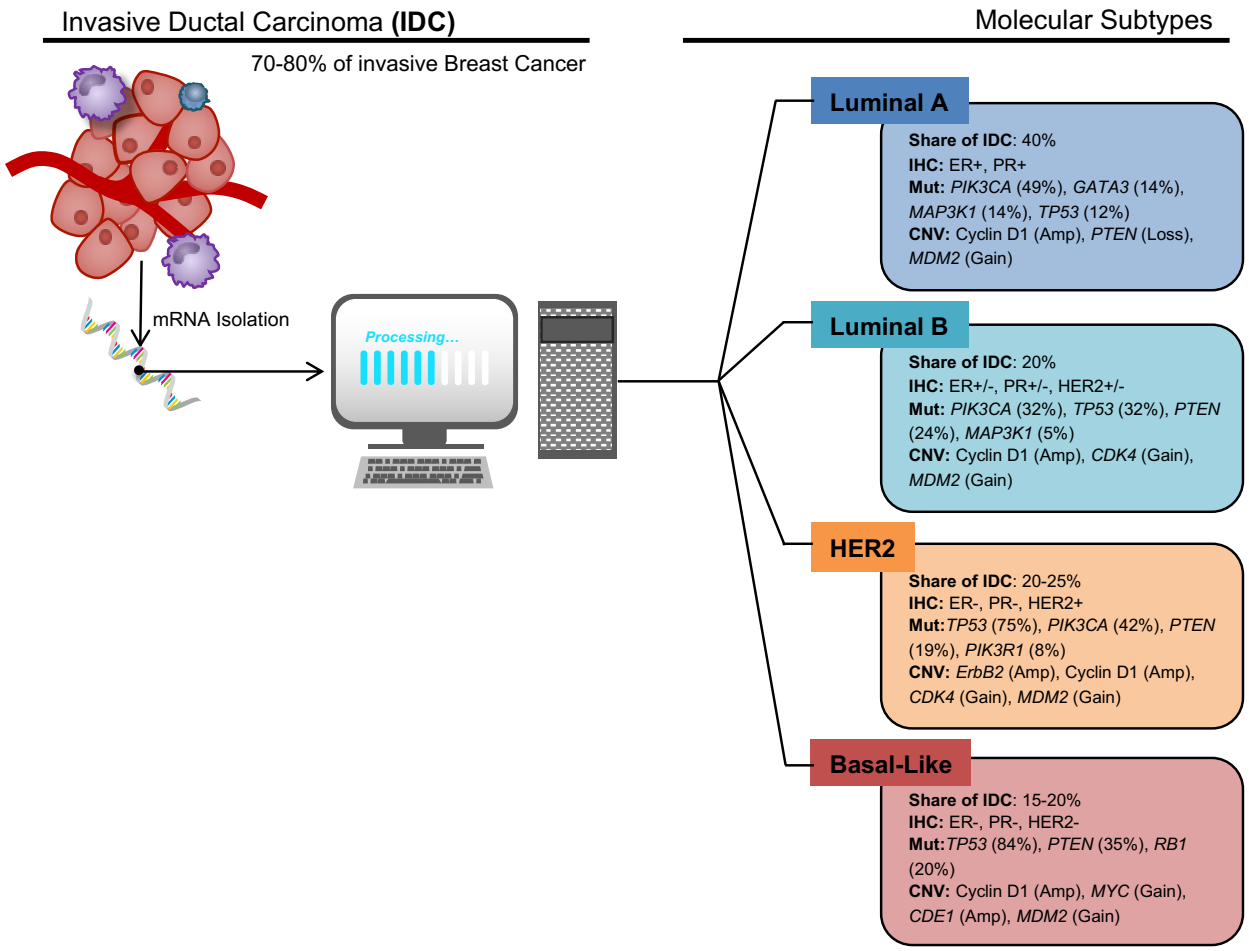


**Figure 2.** Four major molecular subtypes of IDC by gene expression patterns. Invasive ductal carcinoma can be divided into four major molecular subtypes by gene expression patterns. These subtypes aid in the application of therapies and the prediction of disease course. IHC, immunohistochemistry; Mut, mutations; CNV, copy-number variations.

*Big data is an understatement*

Over the course of the last 50 years, a governing principal in integrated circuitry has been Moore's Law. In 1970, the co-founder of Intel, Gordan Moore, predicted the doubling rate of transistors on integrated circuits, like computer processors, will double every two years.[27] There is no better example of the indefatigable march forward of processing speed predicted by Moore than the sequencing the human genome. Although the Human Genome Project cost a total of $2.7 billion and took 13 years, the first working draft of the human genome sequence took 15 months to produce and cost an estimated $300 million.[28] In the 17 years since the completion of the Human Genome Project, the cost of sequencing the human genome is approaching $1,000 and takes less than a week. Interestingly, this reduction in time to sequence a genome fits Moore's Law, with the speed of sequencing doubling 10.4 times in 17 years. The force multiplier that is big data had been brought to biology and spurred on the National Cancer Institute and the National Genome Research Institute announcement of the Cancer Genome Atlas (TCGA) project. Over the course of 10 years, the TCGA spent $375 million dollars

multimodally quantifying over 10,000 tumor samples, generating 20 petabytes of data (one petabyte equals 1024 terabytes).[29]

The largest dataset from the TCGA, breast cancer includes quantification of 1,098 primary breast tumors, 113 paired normal breast samples, and six metastases.[31] These samples were assayed for DNA mutations (whole-exome and whole-genome sequencing), copy number variations, RNA expression (RNA sequencing and microarray), DNA methylation, and protein using the reverse-phase protein array (RPPA). The resulting data provided researchers with a "comprehensive molecular portrait" of breast cancer or rather a previously unknown resolution in the investigation the molecular underpinnings of breast cancer.[30] This portrait paints a picture of breast cancer as highly variable between patients but supports previous literature on categorizations. Luminal A and B breast cancers contain greater than 80% $ER^+/PR^+$ samples, high levels of phosphoinositol-3-kinase (PI3K) pathway activation, and *CCND1* amplification. Luminal B tumors contain 15% $HER2^+$ samples, higher proliferation index compared to Luminal A, and a hypermethylated phenotype.[30] Basal-like tumors have high rates of *TP53* mutations and/or gain in *MDM2*, a negative regulator of p53, resulting in over 95% of samples lacking intact p53 regulation. In addition, BLBC had high rates of *RB1* mutation, highly genomically unstable, and hypomethylated.[30] In contrast, $HER2^+$ tumors had high rates of *TP53*, *CCND1*, and *PI3KCA* mutations, in addition to the characteristic amplification of the *ERRB2* gene locus.[30] In addition to these four major molecular subtypes, additional subtyping strategies have been suggested to include claudin-low, a variant of BLBC with more pronounced stem-like and epithelial-to-mesenchymal (EMT) characteristics.[31] While other groups have identified other TNBC/BLBC subtypes on the basis of predicted drug response by gene expression patterning.[32]

Preclinical and clinical trials for targeted therapies in BLBC have focused on several targets identified by these large datasets. However, many of the targeted approaches have failed due to capacity of BLBC to develop resistance. Investigations into growth factor signaling antagonists, like EGFR pathway or vascular endothelial growth factor receptor (VEGF) signaling demonstrated no clear improvement in survival for patients.[33–35] These trials suggest a capacity for innate and adaptive resistance mechanisms to targeted therapies in BLBC.[35] More recently, Poly-ADP-ribose polymerase (PARP) inhibitors have shown promise in BLBC patients with germline mutations in *BRCA1/2*, leading to a recent FDA approval.[36] PARP inhibitors work by interrupting single-strand DNA repair. In the context of tumors with *BRCA1* mutations, cells are unable to repair single- or double-strand DNA breaks leading to apoptosis, a process also referred to as synthetic lethality.[37] Despite the promise of PARP inhibitors, *BRCA1/2* germline mutations account for 2-5% of total breast cancer incidence and 15-20% of TNBC.[30,38] Trials using PARP inhibitors in TNBC without regard to *BRCA* status had near zero overall response rates.[39] The ability of data to identify potential targeted therapies have directed a precision-medicine initiative by the National Cancer Institute (NCI) to match therapies with underlying

molecular/genetic profiling regardless of the site of origin of the tumors, known as the NCI-MATCH trial. This is a modern molecular take on the Bayesian approaches of diagnosing and treating cancer.

Much like Moore's law and the prediction of current processor speed, the quantifications made by the TCGA is quickly becoming out of date. Emerging from the wake of the TCGA data is the advent of single-cell RNA sequencing and DNA methylation technologies that are able to quantify not only the heterogeneity of cancer cells within a single tumor, but also the interface of cancer cells and stromal cells.[40–42] Previous hypotheses that govern cancer biology, like EMT, a genetic reprogramming incurred as a cancer cell become capable of metastasis, can be investigated on the single-cell level. In one of the first single-cell RNA sequencing projects examining breast cancer, researchers found intratumoral heterogeneity, which was highly subtype dependent.[43] This subtypic difference was also seen in the infiltration of immune cells, with Luminal and HER2+ immune infiltrate principally comprised of B and T cells, while BLBC tumor infiltrates were composed of macrophages and T cells.[43] This intertumoral difference based on breast cancer molecular subtypes in immune infiltrates is intriguing as a therapeutic target and is the basis of the later chapters of this dissertation.

**Anti-tumor immunity and immunotherapy**

If cancer is the betrayal of one's body, immunotherapies function by returning control to the body. The importance of anti-tumor immunity and the promise of immunostimulants against cancer is not a new concept. One of the original pioneers in this field was the bone sarcoma surgeon William Coley. Working in the 1890s at New York Cancer Hospital, later known as Memorial Sloan Kettering Cancer Center, Dr. Coley noticed his patients that suffered from post-operative infections after tumor resections had markedly better survival. This observation led Dr. Coley to inject bacterial cocktails in hopes of stimulating his patients' anti-tumor immunity.[44] This risk of injecting live bacteria into cancer patients, who are already immunosuppressed, led Dr. Coley to further pursue attenuated mixtures of bacterial product and over the course of his career he estimated to have used his therapy on over 1,000 patients. Despite the promise of this work, many of Dr. Coley's peers were highly skeptical and the cancer immunology field went into dormancy with the advent of radiation and chemotherapies.[45] Fast forward to 2013 and cancer immunotherapies were named the Science Breakthrough of the Year from the American Association for the Advancement of Science and the 2018 Nobel Prize in Physiology or Medicine was awarded to James P. Allison and Tasuka Honjo for the unleashing the anti-tumor potential of T cells through immune checkpoint blockade.[46,47]

In the original seminal review of cancer, the Hallmarks of Cancer, Douglas Hanahan and Robert Weinberg did not identify the tendency of cancer to avoid immune destruction, a lapse that was corrected in the follow-on 2011 version of the review.[48,49] In the same year, 2011, the physician scientist Ralph Steinman succumbed to his pancreatic cancer, three days before he was awarded the Nobel Prize in Physiology or Medicine for his

identification of dendritic cells, the first posthumous award in the history of Nobel Prizes.[50] Dr. Steinman's final project was the development of a dendritic-cell-based cancer vaccine to activate his own immune system against his pancreatic cancer. In the interim 100 years between Coley's cocktail to re-emergence of cancer immunotherapy, what changed to make cancer immunology a viable therapeutic option? As Dr. Steinman approach illustrates, the field found the ability to give patients' immune systems a direction against cancer by using T cells. As I complete this dissertation, the announcement of the 2018 Nobel Prize for immune checkpoint blockade (ICB) serves as a poetic bookend for the identification of T cells as important and specific targets for cancer immunotherapy.

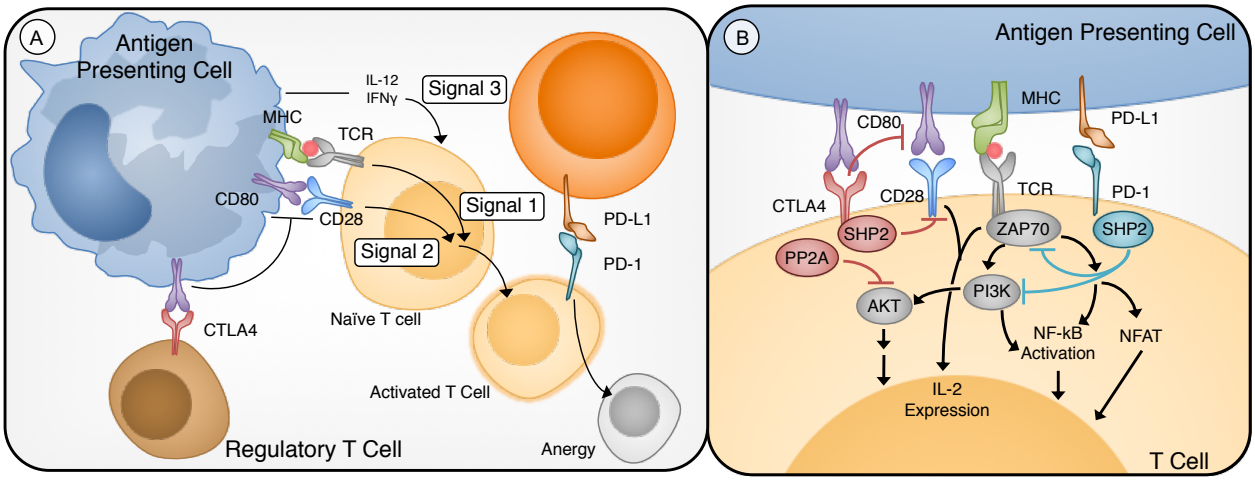


**Figure 3.** T-cell activation and the role of immune checkpoints. **A**. General scheme of T-cell activation Signal 1 is provided by the TCR binding to the antigen presented on MHC. Signal 2 is the costimulation of the T cell by the interaction of CD28 on the T cell with CD80 or CD86. Cytokines act as a Signal 3 that directs the T-cell differentiation. **B**. Mechanistic summary of TCR activation and points of CTLA-4-mediated (red) and PD-1-mediated (blue) inhibition.

*T-cell activation and response*

T-cell activation is a multilayered process involving cell-to-cell communication between antigen-presenting cells (APC) and T lymphocytes (**Figure 3A**). This model of T-cell activation is referred to as the two-signal model with a tertiary signal directing the type of immune activation. The initial signal comes from interactions between the T-cell receptor (TCR) on naïve $CD4^+$ or $CD8^+$ T cells and processed antigenic peptides via the major histocompatibility complex (MHC) on APCs.[51] Subsequently, the intracellular domains of the TCR are phosphorylated by Lck, leading to the recruitment and activation of Zap70, the major relay of TCR activation.[52,53] A network of anti-apoptotic and proliferative signals downstream of Zap70 provide the primary signal for activation of T cells.[54–56] The second signal, or co-stimulatory signal, comes via interaction between co-stimulatory receptors on T cells with their ligands on APCs. The most prominent co-stimulatory signal is the interaction of CD28 on the T cell with CD80 (B7-1) or CD86 (B7-2). This interaction leads to increased PI3K-mediated proliferative signaling and cytokine production.[57,58] A tertiary level of signaling is provided by

cytokines, directing lineage-specific expression of transcription factors and shaping the type of immune response.[59] The activation of T cells is balanced by co-inhibitory receptors on T cells, often upregulated upon TCR activation, which counterbalances co-stimulatory signals.[60] These co-inhibitory signals are referred to as immune checkpoints. Together, the signals from MHC-TCR, CD28-CD80/86, and cytokines determine the magnitude and effectiveness of T-cell responses to infection or cancer.

Immune evasion is recognized as one of the hallmarks of cancer,[49] and therapies which elicit an anti-tumor immune response, either by directly stimulating an immunogenic response or targeting inhibitory pathways, have long been sought after. Tumors can function to subvert adaptive immune activation by decreasing expression of MHC-I molecules to prevent the activating signal or through directing the T cell differentiation or activity.[61–64] The latter is a diverse method involving the interaction of cells in the tumor microenvironment and can function by increasing infiltration of immunosuppressive regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), decreasing antigen presentation, increasing the secretion of suppressive cytokines, or the overexpression of inhibitory ligands for ICB, like programmed death-ligand 1 (PD-L1).[63–65]

*Immune checkpoints: check yourself before you wreck yourself*

CTLA-4 is a member of the CD28 family and is expressed exclusively by T lymphocytes. Upon T-cell activation, CTLA-4 within intracellular granules is translocated to the plasma membrane.[66] This fast translocation allows for CTLA-4-mediated regulation of the amplitude of T-cell response by dampening T-cell activation and priming. CTLA-4 binds to CD80 or CD86 costimulatory molecules, acting as a competitive antagonist with CD28.[67] In addition, CTLA-4 works through protein tyrosine phosphatase 11 and 6 (also known as SHP-1/2) to dampen TCR signal.[67] CTLA-4 plays a role in one arm of peripheral tolerance, an immunological process to prevent self-reactive immune responses, by dampening T effector cell function and increasing immunosuppressive Treg activity.[68] Unlike effector cells, Tregs express CTLA-4 constitutively and CTLA-4 acts as a major mechanism of suppression by Tregs.[69] CTLA-4 on Tregs competes with CD28 on effector T cells for binding with CD80/86 on APCs, thus suppressing T-cell activation. The higher level of CTLA-4 on Tregs also serves to preferentially deplete Tregs in tumors treated with anti-CTLA4 therapies.[70] Increasing the binding affinity the Fc portion of the antibody enhanced the depletion of Tregs and increased response in tumors.[71] Due to the earlier inhibition mediated by CTLA-4, the blockade of CTLA-4 leads to non-specific immune cell activation and is associated with increased treatment-related adverse events.[72] This nonspecific immune activation is underscored by the early death of *Ctla4*-null mice due to widespread lymphoproliferative disease within the peripheral T cell compartment.[73,74] In humans, disrupted expression and localization of CTLA-4 has been linked with type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, and collagen-induced arthritis.[75,76]

PD-1, another member of the CD28 family, acts as a negative regulator of immune response preferentially in peripheral tissues through the interaction with PD-L1 or PD-L2. Part of a negative feedback loop after immune activation, PD-1 and PD-L1/2 play a role in the maintenance of peripheral tolerance and the letdown following immune activation or chronic infections. As a part of peripheral immune suppression, PD-1 is expressed on lymphocytes, monocytes, NK cells, and dendritic cells.[77] Importantly, PD-1 is absent on resting or naïve T cells and is transiently upregulated during the activation process. This upregulation is analogous to the development of T cells in the thymus, with increased PD-1 required for positive and negative selection of immature T cells following TCR activation.[78] Upon activation T cells upregulate the transcription of *PDCD1*, the gene locus for PD-1. The transcriptional regulation of PD-1 is different from CTLA-4, which localizes to the plasma membrane quickly after TCR signaling.[66] PD-L1 is broadly expressed, and the protein is easily to be induced by many cytokines, particularly type 1 and type 2 interferons.[79–82] In contrast to PD-L1, the second ligand, PD-L2, is exclusively expressed on APCs.[81,83,84]

PD-1 suppression acts downstream of TCR activation through the stimulation of protein tyrosine phosphatases SHP1/2. In turn, SHP1/2 dephosphorylates ZAP70 and PI3K, the major targets of the costimulatory CD28 (**Figure 3B**).[85,86] Additionally, there are reports that suggest PD-L1 and transforming-growth factor β (TGFβ) can convert naïve $CD4^+$ T cells into induced Tregs, which potentially provides a long-term immune suppression by PD-1/PD-L1 pathway.[87,88] The loss of PD-1 suppression is associated with antibody-mediated autoimmunity, like glomerulonephritis, dilated cardiomyopathy, and lupus-like autoimmunity.[89–91] Underscoring the importance of PD-1 in peripheral tolerance, single-nucleotide polymorphisms in PD-1 have also been associated with a wide range of autoimmune conditions, including systemic lupus erythematosus, type I diabetes, multiple scleroses, rheumatism, Grave's Disease, and ankylosing spondylitis.[92] This link to antibody-related autoimmunity is likely a direct result of PD-1-mediated suppression of B cell activation.[92]

Sustained expression of PD-1 on T cells is a marker for T-cell exhaustion.[93] Exhausted T cells arise from activated effector T cells that gradually become silenced due to persistent exposure to antigen.[94] Exhausted T cells become progressively dysfunctional due to increased expression of several inhibitory receptors, including PD-1 and CTLA-4, a gradual loss of effector cytokine secretion, and altered cellular metabolism.[94,95] T-cell exhaustion, which is thought to have evolved as a mechanism to promote peripheral tolerance and prevent autoimmunity, is often co-opted by viruses during chronic infection and cancers as a mechanism of immune evasion.[96,97] Effector T cells in the tumor microenvironment become exhausted due to constant exposure to tumor antigens and upregulation of PD-L1 on cancer cells and myeloid cells by oncogenic signaling and inflammatory cytokines.[97–99] It is thought that anti-PD-1 and PD-L1 therapies work in part by reversing or preventing T-cell exhaustion.

Outside of PD-1 and CTLA-4, there are a number of immune stimulatory and suppressive molecules currently under investigation for therapeutic targeting as cancer immunotherapy.[72] Similar to PD-1/PD-L1 and CTLA-4, other inhibitory checkpoints, like LAG-3, TIM-3, BTLA, VISTA, or TIGIT can dampen the anti-tumor immune response by regulating T-cell activity.[100,101] Like PD-1, the negative immune checkpoints, TIM-3, LAG-3, and BTLA are expressed on T cells subsequently to T-cell activation and have been implicated as markers of T-cell exhaustion in tumors.[101–105] TIM-3 also appears to shift the immune responses by negatively regulating Th1 $CD4^+$ T cells and cytotoxic $CD8^+$ T cells, responsible for anti-tumor immune responses.[106] In contrast, VISTA is expressed by both APCs and T cells, with high expression on the suppressive MDSCs and Tregs, functioning in both myeloid cell activation and Treg function.[107] TIGIT is expressed by Tregs, T cells, and NK cells and bind poliovirus receptor on tumor cells or APCs. TIGIT appears to have direct and indirect suppressive effects, by acting as a competitive antagonist to CD226 binding by NK and T cells, but also leading to the recruitment of SHP1/2 and downstream inhibition of AKT signaling in T cells.[108–111] Furthermore, binding of poliovirus receptor to TIGIT on APCs and Tregs increases suppressive activity and release of inhibitory cytokines.[109,112]

In contrast, immunostimulatory checkpoints, like ICOS, OX-40, or CD40L, assist in the activation and maintenance of effector T cells. As a member of the CD28 family, ICOS binds B7H/B7RP-1 and can function in providing the second signal in immune activation.[113] Likewise, OX-40 can act as a costimulatory signal, but the expression is induced upon T-cell activation and leads to the expression of anti-apoptotic factors BCL-2 and BCL-XL, which sustain the proliferative response in T cells.[114,115] OX-40 is also constitutively expressed on Tregs and binding of the receptor to OX40L decreases Treg function.[116] The interaction of CD40 (on APCs) and CD40L on activated T cells promotes a proinflammatory immune responses via the induction of NF-kB signaling.[117] Unlike the inhibitory checkpoints where antibodies are used to physically obstruct interaction by the respective ligand, stimulatory checkpoints can be targeted by ligand-expressing viral particles, recombinant ligand peptides, or agonistic monoclonal antibodies. As these emerging immune checkpoints are further developed, the wide range of mechanisms and targeting strategies will broaden efficacy of immunotherapies for patients by allowing physicians to better select therapies for individual tumors.

*Immune checkpoint inhibitors: results of monotherapy trials*

CTLA-4 and PD-1/PD-L1 are the two most widely studied immune checkpoint pathways with monoclonal antibodies that target these two pathways having been approved by FDA for cancer therapy. The success of these agents has spurred on development of other antibody-based or immunostimulatory drugs to be used as monotherapies or in combinatorial regimes. To date, there has been no comprehensive summary of single-agent ICB therapies. Analysis of these results and factors identified by these trials is the first step in understanding the patterns of response and resistance in patients receiving ICB. We performed a

comprehensive summary of results for trials with at least one arm comparing monotherapy of anti-PD-1 (pembrolizumab, nivolumab), anti-CTLA-4 (ipilimumab), or anti-PD-L1 (atezolizumab). These four antibodies were selected based on trial results in multiple cancer types and accounted for 8,069 patients across 43 clinical trials (**Table 1**). Inclusion criteria for summarized trials are the enrollment of clinical trial into ClinicalTrials.gov database and the measurement of objective response using clinical or radiographic data.

**Table 1**. Comprehensive Summary of clinical trials using single-agent immune checkpoint

| Cancer Type | Agent | Target | Total | Objective Response | TRAE 3+ | Total 3+ AE | Citations |
|---|---|---|---|---|---|---|---|
| Melanoma | Pembrolizumab | PD-1 | 1179 | 24.1-33.4% | 5.9-14% | 19-34% | [118–120] |
| | Nivolumab | PD-1 | 620 | 27.6-43.7% | 11.7-16% | 34-43.5% | [121–123] |
| | Ipilimumab | CTLA-4 | 723 | 9.8-19.05% | 19.6-27% | 33.2-55.6% | [118,121,124–127] |
| NSCLC | Pembrolizumab | PD-1 | 977 | 18.9-44.8% | 9.49-26.6% | Not Reported | [128–131] |
| | Nivolumab | PD-1 | 747 | 12.8-26.1% | 10-18% | 30.7-46% | [123,132–134] |
| | Atezolizumab | PD-L1 | 566 | 13.6-17% | 12-14.8% | 44.4-45% | [135,136] |
| UC | Atezolizumab | PD-L1 | 675 | 13.4-26.3% | 7-16% | 12.6-54.4% | [137–139] |
| | Pembrolizumab | PD-1 | 266 | 21.1% | 15% | 52% | [140] |
| | Nivolumab | PD-1 | 256 | 19.6% | 18 | Not reported | [141] |
| RCC | Nivolumab | PD-1 | 607 | 20.8-27.27% | 11.3-18.7% | Not Reported | [123,142,143] |
| HL | Pembrolizumab | PD-1 | 241 | 65-69%* | 6.7-100% | 21.4% | [144,145] |
| | Nivolumab | PD-1 | 103 | 66-87%* | 31.3-43.5% | 76.3-100% | [146,147] |
| Gastric Carcinoma | Nivolumab | PD-1 | 268 | 11.2% | 10% | 41.5% | [148] |
| HNSC | Pembrolizumab | PD-1 | 236 | 13.3% | 13.1% | 39.4% | [149] |
| HCC | Nivolumab | PD-1 | 214 | 19.6% | 25% | 60.4% | [150] |
| Sarcoma | Pembrolizumab | PD-1 | 80 | 5-18% | 7.1-11.9% | 54.8-59.5% | [151] |
| | Nivolumab | PD-1 | 38 | 5.3% | 28.6% | Not Reported | [152] |
| MSI-H Tumors | Pembrolizumab | PD-1 | 81 | 25-80% | 20.24% | Not Reported | [153] |
| TCA | Pembrolizumab | PD-1 | 40 | 22.5% | Not Reported | 77.5% | [154] |
| Anal Carcinoma | Nivolumab | PD-1 | 35 | 24% | 16% | Not Reported | [155] |
| Breast | Pembrolizumab | PD-1 | 27 | 18.5 | 15.6 | Not Reported | [156] |
| Merkel Cell Carcinoma | Pembrolizumab | PD-1 | 26 | 56% | 15% | 42.3% | [157] |
| CLL | Pembrolizumab | PD-1 | 25 | 12% | 60% | Not Reported | [158] |
| Hematological Malignancies | Ipilimumab | CTLA-4 | 22 | 31.8%* | Not Reported | 86.4% | [159] |
| PMBCL | Pembrolizumab | PD-1 | 17 | 41% | 11% | Not Reported | [160] |

TRAE 3+, Grade 3 or Above Treatment-Related Adverse Event; AE 3+, Total Grade 3 or Above Adverse Event; NSCLC, Non-small cell lung carcinoma; RCC, Renal cell carcinoma; UC, Urothelial carcinoma; HL, Hodgkin's lymphoma; HCC, Hepatocellular carcinoma; PMBCL, Primary mediastinal B-cell Lymphoma; HNSC, Head and neck squamous cell carcinoma; CLL, Chronic lymphocytic leukemia; TCA, Thyroid carcinoma. *Utilized different response criteria between studies

Based on the results of these trials, major predictors of response to ICB are the therapeutic targets and the tumor types. In general, agents targeting PD-1 have higher rates of response with lower toxicity, compared to anti-CTLA-4 antibody, ipilimumab. Evaluation of both the treatment efficacy and treatment-related adverse events (TRAE) of anti-CTLA-4 *versus* anti-PD-1 is limited by the broader use of ipilimumab in combination with other agents that are not included in this summary. Restricting analysis to the 10 single-agent trials in melanoma, objective responses were 9.8-19.1% for ipilimumab compared to 24.1-43.7% for patients on anti-PD-1 therapies.[118,120,121,161–163] Previous work has reported increased TRAEs with anti-CTLA-4 compared to anti-PD-1 therapy, which is seen in comparing 10 trials of melanoma with grade 3 or above TRAE for ipilimumab ranging from 19.6-27% compared to 5.9-16% for anti-PD-1 therapies.[72] Several complications exist comparing multiple trial results across single tumor types. For example, the high variability in objective responses to pembrolizumab (18.9-44.8%) and nivolumab (12.8-26.1%) in non-small cell lung carcinoma (NSCLC) is likely partially a result of different inclusion criteria for patients based on PD-L1 immunohistochemical staining discussed later.[128,130,131,133,134,163–165] Similarly, most of the trials use the criterion for objective response based on the reduction in tumor size greater than 30% as defined by the RECIST criteria.[166] Notably each of the single-agent clinical trials in Hodgkin's Lymphoma utilized different criteria to measure objective response.[145,146,167,168] Likewise, the definition of TRAE varies widely by tumor type with researchers reporting all TRAEs or only TRAEs that affect a certain percentage of the ICB-treated patients.

Tumor type is a predictive correlate of response with higher responses for both anti-PD-1 and anti-CTLA-4 agents in immunogenic tumors and lymphomas. Of note, in advanced, refractory Hodgkin's Lymphoma responses ranged between 65-87% across both pembrolizumab and nivolumab.[145,146,167,168] High response rates were also seen by tumors with high microsatellite instability (MSI-H, 25-80%, pembrolizumab)[169], Merkel cell carcinoma (56%, pembrolizumab)[170], and primary mediastinal B-cell lymphoma (PMBCL, 41%, pembrolizumab).[160] Although beyond the scope of the inclusion criteria for the table, early phase trials included other solid tumors that did not demonstrate response to ICB, notably pancreatic cancer[171], castration-resistant prostate cancer[163], colorectal cancer[163], gastric cancer[172], and breast cancer.[172]

**Predictors of resistance and response for ICB therapy**

The diverse clinical response of patients across multiple tumor types to ICB highlights the need to understand mechanisms that drive response and resistance. A variety of factors that influence immune response are directly or indirectly influenced by the tumor, providing permissive conditions for tumor growth and progression. For example, chronic inflammation can drive tumor growth by providing growth factors and preventing innate or adaptive anti-tumor immune responses.[173,174] In turn, the tumor can produce chemotactic agents and other factors that maintain the inflammation in a positive feedback loop.[173,174] The inflammation in cancer is influenced by systemic factors, like age, adiposity, genetics, or the gut microbiome. Like the factors that impact

chronic inflammation and cancer, predictors of response to ICB can be divided into three categories by location: tumor cell intrinsic, tumor microenvironment, and systemic factors. Despite the tidy organization, many of the predictors of response interact at multiple levels and ultimately translate to modulating total infiltration and function of immune cells in the tumor before or after ICB therapy.

*Tumor-intrinsic factors that predict response to ICB*

Whole-exome sequencing has revealed the mutational load of tumors as a major predictor of ICB response.[153,165,175–177] Underlying these observations is the hypothesis that somatic mutations lead to tumor-specific neoantigens that can overcome self-tolerance (**Figure 4A**). However, high mutational load is not necessary for ICB response as in the case of clear cell renal cell carcinoma (ccRCC) with greater than 20% objective response rates and a relatively low mutational load of 1 mutation/megabase, compared to the 10 to 400 mutations/megabase in melanoma or NSCLC.[163,178–180] Mutator phenotypes, like those with familial Lynch Syndrome or other mismatch repair deficient (dMMR) tumors, have objective response rates ranging from 25-80% for anti-PD-1 therapies.[153,169] These high response rates led to the accelerated approval in May of 2017 of pembrolizumab for unresectable or metastatic dMMR or microsatelliteinstability-high (MSI-H) cancers regardless of their tissues of origin, a first such approval from the FDA that is based on molecular trait rather than cancer type. In NSCLC, increased C-to-A nucleotide transversions and decreased C-to-T transversions were found to be indicative of smoking status and may serve as a marker of response to ICB.[165] Like Lynch Syndrome with mutations in DNA repair genes, most commonly *MLH1* and *MSH2*, mutations in *BRCA2, POLD1,* and *POLE* have also been identified in association with response to anti-PD-1 therapies.[165,181] *BRCA2* is an adaptor for homologous recombination, while *POLD1* and *POLE* code for subunits of DNA polymerases that synthesize and repair DNA. Mutations in these three genes may also increase total mutational load in tumors. In a similar vein, a current open-label phase II clinical trial involving breast cancer patients with *BRCA1/2* germline mutations (NCT0302503) is being conducted and will assess if mutator phenotypes are a predictor of response outside of immunogenic tumors. Preclinical transgenic murine models of *BRCA1*-deficient breast tumors demonstrated combinatorial therapy of anti-CTLA-4 and anti-PD-1 with the DNA crosslinking agent, cisplatin, enhanced survival.[182] Unlike mutations, chromosomal copy number alterations and aneuploidy has been reported to be inversely associated with response to ICB due to the possible loss of immune mediators.[183,184]

In the context of tumor-specific mutations, both quantity and quality play a role. Evaluating neoantigens on the basis of likelihood of presentation and T cell recognition, termed neoantigen fitness, has been found to predict increased immunotherapy response.[185] Parsing neoantigen fitness is a basis for investigations into personalized neoantigen-based vaccines.[185–189] In combination with anti-PD-1 therapies, personalized-vaccine priming appears to work synergistically in promoting anti-tumor immune response.[186,189] Anti-PD-1 therapy

has been reported to contract the mutational load and cancer cell clonality in complete and partial responders; whereas nonresponders tend to have persistent mutational load and cancer cell clonality.[190] Beyond DNA repair-related genes, mutations in *SERPINB3* and *SERPINB4* mutations, a class of peptidase inhibitors, have been reported in responders to anti-CTLA-4 and anti-PD-1.[190,191] Recent data from anti-PD-1 therapy in ccRCC found loss-of-function mutations in *PBRM1*, a chromatin remodeling subunit, to be associated with response and better overall survival.[180] Conversely, mutations in *JAK1* and *JAK2* kinases, major relays in interferon γ (IFN-γ) signaling in melanoma and epidermal growth factor receptor (*EGFR*) in NSCLC have been identified in progressors on ICB therapy.[192,193]

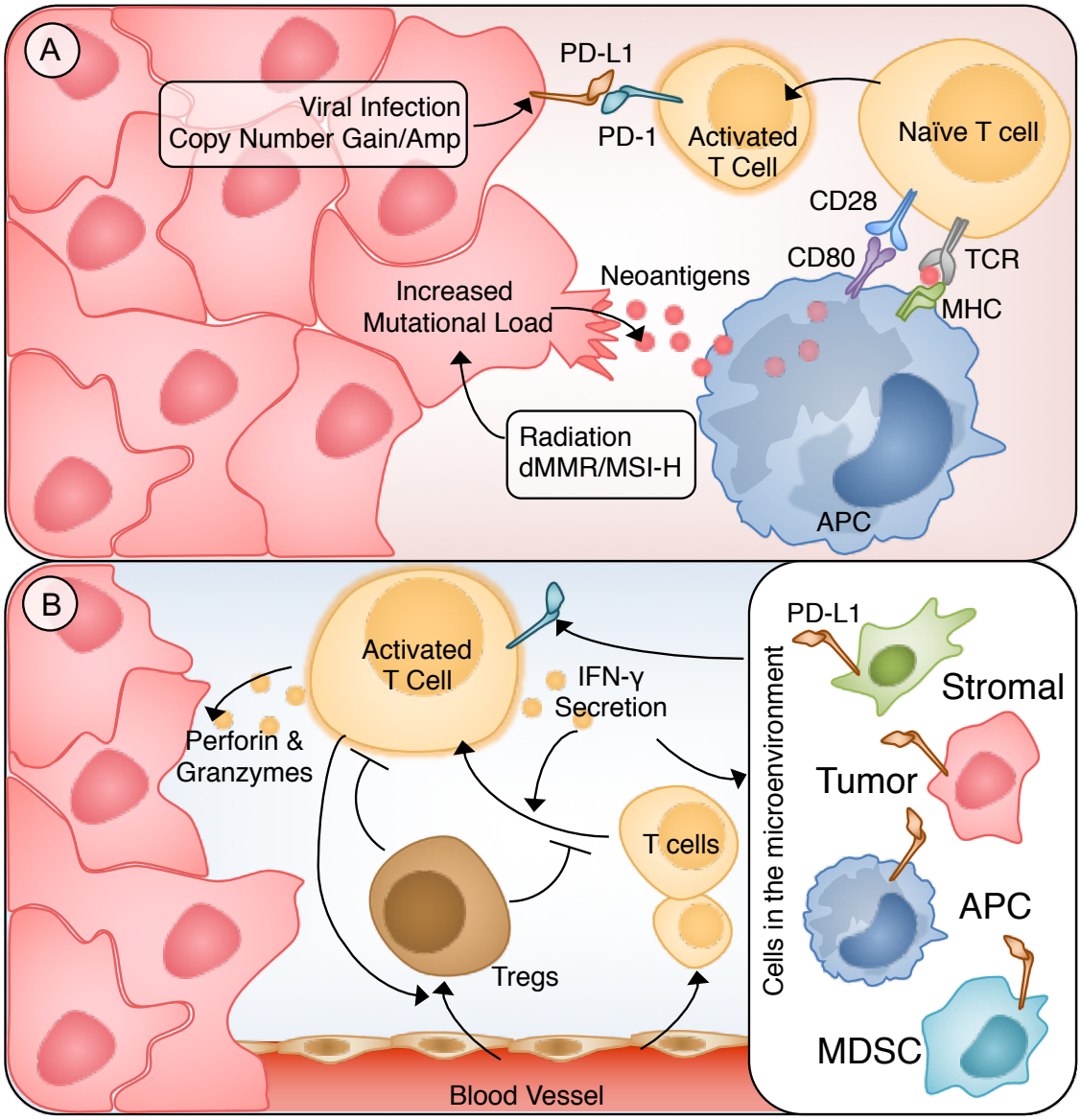


**Figure 4**. Factors that influence response of tumors to immune checkpoint blockade. **A**. Tumor intrinsic markers of response to therapy, focusing on increased mutation and expression of PD-L1by tumor cells. **B**. Tumor microenvironmental factors that influence response. Increased activated T cells lead to IFNγ that drives PD-L1 expression on other cells. Increased Tregs can suppress anti-tumor immune response.

In the context of anti-PD-1 therapies, PD-L1 immunohistochemical staining and gene expression in tumor cells have been proposed to predict response to PD-1 therapy in melanoma, NSCLC, ccRCC, and Hodgkin's lymphoma (**Figure 4A**).[134,146,194–200] This corresponds to the proposed mechanisms of anti-PD-1 therapies, in which cancers with high levels of PD-L1 expression are suppressing anti-tumor immunity and the blockade overcomes the PD-1/PD-L1-mediated immune suppression. The use of PD-L1 staining in NSCLC using a cut-off of 5-10% positive cells has mixed benefit.[132,134,201,202] Using a higher cut-off of 50% PD-L1-positive cells in the Keynote-001 study of pembrolizumab in NSCLC, PD-L1$^+$ patients had an objective response rate of 45%, compared to 19.4% across all patients.[130] This predictive finding led to FDA approval of PD-L1 immunohistology as a companion diagnostic for pembrolizumab in 2015. Within melanoma, the staining for PD-L1 in greater than 5% of cells is more consistently associated with increased objective responses to anti-PD-1 therapy based on a meta-analysis of 11 melanoma trials.[202] Interestingly, PD-L1 expression is not necessarily associated with regional or total lymphocytic infiltration.[200] These collective disparities may be a result of a greater association of immune infiltration with tumors containing PD-L1$^+$ immune cells compared to PD-L1$^+$ tumor cells alone.[203]

Immune cells may drive PD-L1 expression on tumor cells through interferon signaling.[79–82] In melanoma, IFN-γ-mediated PD-L1 expression requires downstream activation of JAK/STAT, highlighting the role of previously mentioned mutations in *JAK1/ JAK2* in resistance to anti-PD-1 therapies.[81,192] Immunogenic etiologies of increased PD-L1 expression has been reported in Merkel cell carcinoma with infection by Merkel cell polyomavirus or induced by the combinatorial ICB and oncolytic virus therapy.[170,204] Despite an association of Hodgkin's Lymphoma with the Epstein-Barr virus, no association between infected patients *versus* uninfected patients and PD-L1 expression has been reported.[146] Regardless of infection status, the Reed-Sternberg tumor cells of Hodgkin's Lymphoma contain copy number gain or amplification of PD-L1 and PD-L2 genetic loci, *CD27*4 and *PDCD1LG*.[146] This may contribute to the up to 87% objective response rates reported in Hodgkin's Lymphoma.[146] Interestingly, elevated PD-L1 expression on melanoma cells has been associated with poor response to combinatorial anti-CTLA-4 and radiation therapy, supporting the complementary, yet distinct roles of CTLA-4 and PD-1 in immune regulation.[205]

*Microenvironmental factors that predict response to ICB*

A product of the immune-tumor microenvironment, immune response gene expression patterns have been linked to improved responses in both anti-PD-1 and anti-CTLA-4 therapies (**Figure 4B**). Common among both of these therapies is activated CD8$^+$ T cell signature and IFN-γ signaling. Principally, increased response has been associated with elevated expression of *CD8A*, *IFNG*, *PRF1*, *GZMA*, and *GZMB*.[192,195,199,206,207] The interface of tumor cells and immune cells plays a role in the activation of immune cells with mutations or copy number loses in both type 1 and type 2 interferons signaling modules associated with decrease response or

resistance to ICB.[192,207] Interestingly, immunosuppressive pathways, like IL-10, have had mixed observations, with *IL10* expression reported lower in responders before the induction of pembrolizumab.[181] Conversely, IL-10 is increased in PD-L1$^+$ melanomas[195] and copy number loss in the α subunit of the IL-10 receptor is has been seen in anti-CTLA-4 nonresponders.[207]

Gene expression signatures are indirect measures of immune cell infiltration and activation in the tumor microenvironment. A major focus of prognostic forecasting for ICB response has been increased T cell presence at the invasive margin or intratumorally.[184,198,208] More specifically, increase in activated CD8$^+$ T cells and decrease in immunosuppressive Tregs have been reported in responders to ICB (**Figure 4B**).[166,178] A hallmark of Lynch Syndrome-associated hereditary nonpolyposis colorectal cancer and MSI-H sporadic colorectal cancer is lymphocytic infiltration, which may partially account for the increased response rates to ICB in these tumors.[209,210] Additionally, MSI-H tumors have been associated with increased cytotoxic CD8$^+$ T lymphocytes and increased IFN-γ-secreting Th1 CD4$^+$ cells compared to microsatellite-stable tumors.[211] Work from Roh, *et al.* suggests that not only quantity of lymphocytic infiltration predicts better response, but also expanded TCR clonal populations before the induction of anti-PD-1 therapy are predictive.[184] Similarly, after the induction of anti-PD-1 therapy, clonal expansion of T lymphocytes was seen exclusively in patients with objective responses.[190] The clonal expansion of T cells, however, does not seem to be associated with the response to anti-CTLA-4.[184]

These works have led to attempts using combinatorial approaches to stimulate or prime the tumor microenvironment to increase T cell infiltration and activity. Notably investigations into using ICB in combination with vaccines,[212] anti-angiogenic molecules,[213] oncolytic viruses,[204] radiation,[205,214,215] suppressors of TGFβ,[216] and DNA-methylation inhibitor have been reported.[217] The use of combinatorial radiation and ICB seems to function at multiple levels, by increasing lymphocytic infiltration, removing suppressive cells, and priming the microenvironment while also increasing mutational load.[205,214,215] Dysregulation following CD8$^+$ T cell infiltration may contribute to resistances to ICB therapies by increasing PD-L1 expression by tumor cells, increasing Treg infiltration, and upregulating the suppressive indoleamine-2,3-dioxygenase (IDO) metabolite within the tumor microenvironment.[81,218] IDO itself may play a role in the resistance to anti-CTLA-4 therapy with clinical trials underway using IDO inhibitors with ICB.[219,220]

Despite the focus on CD8$^+$ T cell infiltration, response has been associated with a number of other immune cells. With PD-L1/TGFβ dual inhibition, increased infiltration of CD8$^+$ T cells, natural killer cells, dendritic cells, and M1 macrophages have been seen in responders.[190,216] Although M1 macrophages have been observed to be increased with PD-L1/TGFβ dual inhibition, total macrophages may be decreased in responders to ICB monotherapy.[190,216] Macrophage polarization may play a larger role, with M1 polarization associated with PD-

L1 expression and M2 polarization associated with PD-L1 and PD-L2 expression.[221] Beyond myeloid cells, PD-L1 staining in tumor cells and immune cells has been significantly associated with presence of $CD20^+$ B cells and lymphoid aggregates.[200] Interestingly, PD-L1 staining on immune cells and tumor cells may have an additive effect in the percentage of intratumoral PD-$1^+$ CTLA-$4^+$ $CD4^+$ and $CD8^+$ T cells.[203] Highlighting the difference in PD-L1 expression by tumor *versus* immune cells, limited evidence indicates that PD-L1 from immune cells is correlated with increased TIM-$3^+$ $CD8^+$ T cells in NSCLC.[203]

*Systemic factors that predict response to ICB*

Systemically, there has been a focus on identify biomarkers of therapeutic response to ICB. Several groups have tested the idea of using circulating immune cells as a metric for response. In peripheral blood, increased total $CD4^+$ and $CD8^+$ lymphocyte count and clonal expansion of specific T lymphocyte have been reported in responders to ICB therapy.[222–225] A more detailed examination of T cell dynamics in peripheral blood has demonstrated an increase at baseline in IL-$4^+$, $GzmB^+$, IFN$\gamma^+$, and GM-$CSF^+$ $CD4^+$ T cells and CTLA-$4^+$, $GzmB^+$, and IL-$13^+$ $CD8^+$ T cells among responders to anti-PD-1 therapy (**Figure 5**).[226] After the start of ICB therapy, responders had an increase in $GzmB^+$ and CTLA-$4^+$ $CD8^+$ T cells and PD-$1^+$, IL-$4^+$, IFN$\gamma^+$, IL-$17A^+$, and $GzmB^+$ $CD4^+$ T cells.[205,226] A greater percentage of exhausted PD-$1^+$ $EOMES^+$ $CD8^+$ T cells in the peripheral blood was associated with decreased response to anti-CTLA-4 and radiation combinatorial therapy.[205] Potential immune cell markers of response are not exclusive to T lymphocytes with increases in myeloid populations being reported in responders.[224,226] Specifically, increases $CD14^+$ $CD16^-$ HLA-$DR^{hi}$ monocytes and eosinophils have been reported in responders after the induction of anti-PD-1 and anti-CTLA-4 therapies, respectively.[224,226]

Human leukocyte antigens (HLA) are a series of genes that encode the MHC that activate and regulate immune response. HLA-typing represents a predictive factor at all three levels: expression on tumors cells, expression on immune cells in the microenvironment, and diversity of HLA molecules based on the genetics of an individual (**Figure 5**). In a large analysis of patients, diversity of the both MHC-I (HLA-A/B/C), found on most nucleated cells in the body, and MHC-II (HLA-$D_n$), found on APCs, have been associated with increased response to ICB.[227] Homozygosity of *HLA-B*, *HLA-A*, *HLA-DP*, and *HLA-DPB* alleles was associated with poorer overall survival in patients receiving ICB therapy.[227] Specific alleles, like HLA-B62 was found to increase the risk of death on ICB, while *HLA-B44* had significantly better overall survival.[227] Inoue, *et al* found increased level of *HLA-A* in pre-treatment resections of patients with objective responses to nivolumab.[199] Likewise, in the MHC-II family, HLA-DR IHC staining in melanoma patients receiving anti-PD-1 or anti-PD-L1 was associated with improved therapeutic response and infiltration of $CD4^+$/$CD8^+$ T cells. Similar to response, TRAE from ICB has been linked with HLA-typing. An early report of anti-CTLA4 trial data examining *HLA-A*0201$^+$ versus HLA-A*0201$^-$* patients found no difference in response, but rather a significant elevation of

immune-related adverse events in the *HLA-A*0201*$^{+}$ individuals.[228] The implication of off-target immune activation based on HLA genotype has also been seen in checkpoint-induced fulminant myocarditis and an association with *HLA-DQ7*.[229] The off-target immune activation, although higher in anti-CTLA-4 therapies, is central to the debate of using ICB in patients with a history of autoimmunity or HLA-type predisposition.[230]

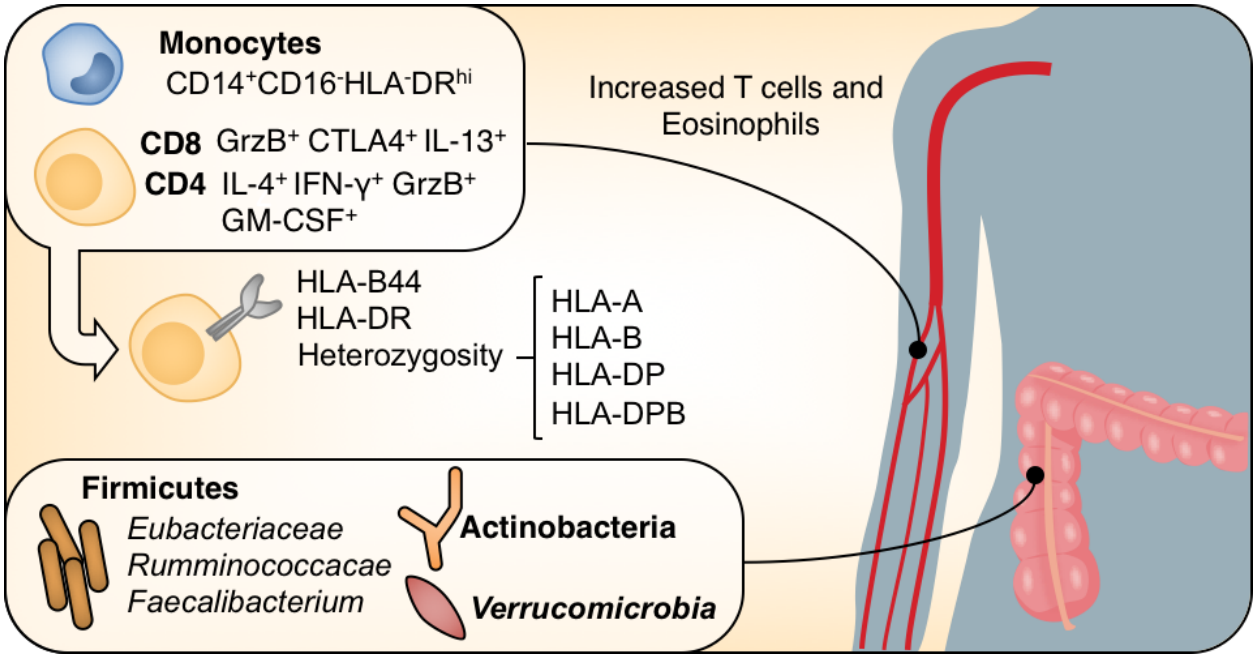


**Figure 5**. Systemic factors that influence response to immune checkpoint. Factors can be divided into immune cells in the peripheral blood associated with better responses, HLA genotypes, and gut microbiome. Increased activated CD4 and CD8 T cells in the peripheral blood have been reported in responders. In addition, HLA diversity and HLA-B44/HLA-DR have been associated with better response. Recent reports of gut microbiome influences on ICB have made it an emerging predictive correlate - for immunotherapy.

Another emerging predictor of response to ICB is the gut microbiome (**Figure 5**). Use of antibiotics has been associated with a reduction in efficacy of ICB in mice and humans.[231,232] Responders receiving PD-1 blockade have increased species in the *Firmicutes*, *Verrucomicrobia*, and *Actinobacteria* phyla.[231,233,234] *Firmicutes*, gram-positive endospore-producing bacteria, form a large portion of the gut microbiome diversity. Within the *Firmicutes* phylum, responders have been observed to have elevated *Eubacteriaceae*, *Rumminococcacae*, and *Faecalibacterium* families in the *Clostridiales* order.[201,203] Additional reports on *Bacteriodetes*, a phylum consisting of gram-negative rods, vary and have been seen elevated in both responders and nonresponders of ICB, which may be a product of different collection time points of post- *versus* pre-treatment, respectively.[201,203] Initial preclinical assessment of supplementing *Firmicutes*, *Verrucomicrobia*, or *Bacteriodetes* species increased T lymphocytic infiltration and response to ICB in murine models.[231–233] Similarly, oral gavage of fecal matter from human responders *versus* nonresponders in murine melanoma tumor models led to differential melanoma growth that recapitulate response to ICB in human patients.[231,233–235] Like HLA typing, the gut microbiome has been associated with TRAE. The use of ICB, more pronounced in anti-CTLA4 therapy, can lead to colitis with mixed immune cell

infiltration consisting of neutrophils, eosinophils, lymphocytes and plasma cells within the lamina propria. This reactive colitis is also accompanied by elevated humoral responses to enteric flora and a decrease in *Bacteriodetes* species.[236,237] Similar observations of immunoglobulin response and decreasing *Bacteriodetes* species have also been reported in inflammatory bowel disease.[238] The underlying mechanistic contribution of the gut microbiome to response to ICB is unknown and may be indirectly contributing to ICB response, modulating the metabolic profile of immune cells, or more directly acting as shared antigens between the tumor cells and gut microflora.[239]

**Purpose of the study**

Data analysis represents the microscope of the twenty-first century. Offering the opportunity to examine biological specimens at a brand-new level of resolution than previously imagined. Management and practical application of big data is a focus of cancer research and biology as a whole. Situated between the bench and the bedside, data drives the development of diagnostic heuristics, aids in the identification of horses and zebras alike, but also drives basic scientific inquiry. Collectively, this work seeks to demonstrate the power of data in these different facets.

RPPA are a medium-to-high-throughput approach to quantify proteins using an antibody-based dot blot. Using protein-based quantifications are particularly advantageous compared to mRNA surrogates because they can act as putative targets for therapies. Within breast cancer, targeted therapy for BLBC represents a significant unmet clinical need. Although BLBC/TNBC tumors respond well initially to therapy, the risk of recurrence for these patients is profoundly increased compared to other clinical or molecular subtypes. In this vein, the focus of chapter 2 is the identification and manipulation of molecular targets in BLBC. With the success of PARP inhibitors for *BRCA1/2*-mutant BLBC/TNBC, modulation of DNA repair has particular interest. Demonstrating the bedside-to-bench approach that data enables, chapter 2 focusing on modulating DNA mismatch repair to enhance anti-tumor immunity and efficacy of immunotherapies. Unlike the inhibition of growth factor signaling in BLBC, immunotherapies offer the opportunity for robust, long-term therapeutic responses. The major hurdle for effective immune therapy in breast cancer is the identification of patients who would benefit and the development of adjuvants that might prime the tumors to be more immunogenic.

Unlike other cancers, such as lung cancers or melanoma, invasive ductal carcinoma is immunogenically weak. Both resident and migrating immune cells have been reported to participate in tumorigenesis and progression. However, clinical trials investigating the efficacy of immunotherapies in breast cancer have had mixed-to-poor results. A theory to explain the lackluster results is that breast tumors and other solid epithelial tumors may have tumor intrinsic or microenvironmental factor that contributes to evasion from anti-tumor immunity. To this end, chapter 3 contains a characterization of a subset of immunosuppressive Tregs from renal, breast, liver,

lung, and colorectal cancers that have enhanced suppressive activity. Upon infiltration into tumors, these Tregs upregulate a GPI-anchored gene, *CD177*, with an accompanying increase in a diverse suppressive genetic program. Characterizing the heterogeneity of tumor-infiltrating Tregs can lead to the identification of novel therapy targets and better predictions of clinical course.

To illustrate a similar approach to the analysis of heterogeneity and predictive capabilities of machine-learning, chapter 4 focusing on Cutaneous T cell lymphomas (CTCL). CTCL is thought to be a malignant proliferation of skin-homing $CD4^+$ T cells due to antigen stimulation. As such, CTCL is characterized by a clonal expansion of central and peripheral memory $CD4^+$ T cells. Using single-cell sequencing of mRNA, we developed a more complex model featuring distinct transcriptomic states within the clonal tumor. Moreover, the transcriptional heterogeneity within the malignant $CD4^+$ T cells was used to make a predictive classification for CTCL disease stage. This work offers insight into the heterogeneity of CTCL, providing better understanding of the transcriptomic diversities within a clonal tumor, which can predict tumor stage and thereby offer guidance of therapy.

**Specific aims**

1. Evaluate the targeting of mismatch DNA repair for therapy in aggressive BLBC.
2. Investigate heterogeneity of suppressive Tregs in the tumor microenvironment.
3. Utilize single-cell RNA sequencing to predict clinical progression of cutaneous T cell lymphoma

## CHAPTER 2
TARGETING DNA MISMATCH REPAIR IN AGGRESSIVE BASAL-LIKE BREAST CANCER

### Rationale

Basal-like breast cancer (BLBC) is an aggressive molecular subtype of breast cancer, accounting for 15% of invasive ductal carcinomas.[240] These tumors are characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), or human epidermal growth factor receptor 2 (HER2), a clinical designation also known as triple-negative breast cancer (TNBC).[20,240,241] Although used synonymously, BLBC can be clinically differentiated from TNBC with the additional immunohistochemical staining for basal cytokeratins 5/6 and account for 70-80% of TNBC.[240,242] Unlike ER$^+$ or HER2$^+$ breast cancers, BLBC do not have approved targeted therapeutic options with notable clinical trial failures in targeting angiogenesis and epidermal growth factor receptor 2 (EGFR) that demonstrated no improvement in overall survival.[243] The absence of approved targeted therapy and the variable responses to chemotherapy in BLBC patients represents a vital and unmet clinical need.

BLBC is also defined by high levels of genomic instability, with p53 pathway alterations seen in 85-95% of tumors.[240] This instability in BLBC is compounded by the characteristic loss in chromosome 5q, consisting of DNA repair genes, *RAD17*, *RAD50*, *MSH3*, and X*RCC4*.[240,244] In addition,15-20% of BLBC possess mutations in *BRCA1* or *BRCA2,* which function in double-stranded DNA repair.[240,245–247] Half of the 15-20% of BLBC tumors with *BRCA1* or *BRCA2* are a result of germline mutations and are associated with a 70-80% lifetime risk of breast or ovarian cancer to individuals.[248,249] Recent success of clinical trials using Poly-ADP-ribose polymerase (PARP) inhibitors, a single-strand DNA repair sensor, in germline *BRCA*-mutated ovarian and breast cancers have led to Food and Drug Administration approval.[250,251] PARP inhibitors are thought to work in *BRCA*-mutant cancers by inhibiting single-stranded DNA repair. In the context of *BRCA* mutations with deficient double-stranded DNA repair, the further inhibition leads to cell death, termed synthetic lethality. Despite the positive successes of PARP inhibitors in trials, the translation of PARP therapy to the majority of BLBC and ovarian cancers with intact *BRCA1/2* have been unsuccessful.[39,252–254]

The success of PARP inhibitors for a subset of BLBC/TNBC patients has motivated the field to investigate the potential of other DNA repair-based targeted therapy. Mismatch DNA repair has traditionally been thought of as a tumor suppressor, with germline mutations identified in the autosomal dominant Lynch Syndrome.[255] Closely associated with hereditary non-polyposis colorectal cancer, Lynch Syndrome also increases the risk in the development of cancers along the gastrointestinal tract, endometrial cancer, and ovarian cancer. Defective DNA mismatch repair leads to increase in mutations, both insertions/deletions and single-nucleotide variations, in repetitive sequences, like DNA microsatellites.[256] This increased mutational rate is thought to

underlie the stark 25-80% therapeutic response rate of anti-PD-1 immune checkpoint therapy in patients with Lynch-Syndrome-associated tumors.[257–259] The result of this phase 2 trial led to the unprecedented approval of anti-PD-1 therapies for metastatic mismatch-repair-deficient or microsatellite instable cancers, irrespective of the site of origin. Beyond Lynch Syndrome tumors, variable efficacy of mismatch DNA repair has been implicated in differential mutations in in late-replicating DNA regions, driving mutational abundance in heterochromatin that is seen in a number of cancers.[260]

Here, we identify at MSH2 and MSH6 protein level as poor prognosis markers for the aggressive BLBC. Within BLBC, MSH2 is associated with decreased genetic markers of immune infiltration and activation. The removal of *Msh2* in syngeneic tumor models of BLBC increased efficacy of anti-PD-1 monoclonal therapy. Taken together, we provide evidence of the viability of targeting DNA mismatch repair as a potential adjuvant to improve immunotherapy responses in breast cancer.

## Materials and methods

*Reverse-phase protein analysis*

RPPA data was downloaded from the TCPA Portal located at http://tcpaportal.org/.[261] Data was processed as previously described and clinical data was attached by Patient/Sample ID.[262] Protein-based survival analysis was performed using the previously-developed TRGAted R Shiny application, code and processed data for all TCGA cohorts are located at https://github.com/ncborcherding/TRGAted.[262] Molecular-subtype designation was based on the TCGA Analysis Working Group pipeline, fitting PAM50 subtypes based on RNA-seq data. Protein *versus* mRNA correlations was performed using log2(x+1) mRNA quantification downloaded from the cBioPortal.[263,264] Correlations for both protein-mRNA and protein-protein comparisons utilized the rank-based Spearman method.

*Differential gene and pathway analysis*

Gene-level HTSeq count data was downloaded from the UCSC Xena Browser at http://xena.ucsc.edu/. Using the DESeq2 R package (v1.16.1), count data were separated in half by MSH2 RPPA quantification and converted into negative binomial distributions. A parametric model was fitted for the data and significance and log2-fold change was computed using Wald testing. P-values for the differential analysis was corrected for multiple hypothesis testing using the Benjamini and Hochberg method.[265] Significance threshold was set at adjust P-value $\leq 0.05$ and log2-fold change $\geq |0.5|$. A total of 2,094 genes met the criteria, with 650 genes significantly increased in MSH2-high BLBC samples and 1,444 genes significantly decreased. For visualization of the expression of specific genes by MSH2 grouping, the expression matrix was regularized log transformed. Ingenuity Pathway Analysis (Qiagen, Hilden, Germany) was performed using the above significance cut-off.

DESeq-based log2-fold change was incorporated with the bias-corrected Z-score for upstream regulator analysis in order to visualized activation relative to mRNA fold-change.

*RNA-immune cell estimates*

Immune and stromal ESTIMATE scores for all breast cancer samples in the TCGA were downloaded from MD Anderson Department of Bioinformatics and Computational Biology (bioinformatics.mdanderson.org).[266] BLBC samples with MSH2 protein quantifications were extracted and separated by mean MSH2 value. In addition to general immune cell difference, log2-RSEM mRNA values for BLBC samples were analyzed using the CIBERSORT support vector machine algorithm, in order to estimate the contribution of individual immune lineages.[267] Absolute estimates of immune cell populations were combined for total immune cell lineages and then were then separated by mean MSH2 protein values.

*Cell culture and Msh2 knockout cell lines*

Fresh 4T1 (CRL-2539), PY8119 (CRL-3278), and HEK293T cells were ordered from the American Type Culture Collection. Cells were maintained in RPMI media (ThermoFisher, Waltham, MA) supplemented with 10% fetal bovine serum (Sigma-Aldrich, St. Louis, MO) and 100 U/ml penicillin-streptomycin. Guide RNAs online tool from MIT/BROAD Institute (http://crispr.mit.edu/). The guide RNA for *Msh2* was 5'-CACCGGCGCCGTGTAAAAGTCGCCG-3'. LentiCRISPRv2 (Plasmid #52961, Addgene, Cambridge, MA) system was ligated using Roche repaid ligation kit ( Roche Holdings, Basel, Switzerland).[268] The HEK293T cells transfected with 6μg of the lentiCRISPR-gRNA plasmid, 3μg of pMD2-VSVG plasmid and 3μg of the ΔCMV using GeneTran III (Biomiga, Roselle, CA). The lentiviral stock was then harvested at 72 hours days and centrifuged at 3000 rpm for 10 minutes at 4˚C, removing the cell debris. The supernatant was filtered through 0.45μm filter, aliquoted, and stored at -80°C.

*Immunoblotting*

Cells were lysed in cell lysis buffer [50mM Tris-HCl pH 7.5, 1mM EDTA, 1mM EGTA, 1% Triton-X100 , 100mM KCl, 50mM NaF, 10mM, 1mM Na3VO4, 100nM aprotinin, 1uM bestatin, 1mM p-Amidinophenylmethanesulfonyl fluoride (PMSF)]. Protein (30μg) were separated by SDS-PAGE and transferred onto PVDF membrane (Bio-Rad, Hercules, CA). Membranes were immunoblotted with the anti-MSH2 and anti-β-actin (Cell Signaling Technology, Danvers, MA). Immunoreactive bands were detected using a chemiluminescence based detection system (Bio-Rad).

*In vitro assays using murine cell lines*

For transwell migration assay, $5x10^4$ of 4T1 and PY8119 cells in serum-free RPMI media were added to the cell culture insert with 8 μM pores (ThermoFisher, Waltham, MA) on a 24-well plate. Complete RPMI media, with

10% fetal bovine serum, was added to the bottom well. The cells were incubated at 37°C with 5% $CO_2$ for 8 hours. The insert was rinsed with distilled water and then the cells on top of the insert membranes were removed using a sterile cotton tip applicator. The insert membranes were stained with 0.4% crystal violet for 30 minutes at room temperature, then rinsed with distilled water and dried. *In vitro* transwell migration was measured by counting the cells on the insert membrane. For transwell invasion assay, cell culture insert membranes were coated with 10% matrigel (Corning Life Sciences, Corning, NY) in serum-free media and the cells were incubated for 24 hours at 37°C with 5% $CO_2$. Proliferation assays were performed using $5x10^4$ cells in 6-well plates with complete RPMI media, then incubated at 37°C with 5% $CO_2$ for 48, 72, and 96 hours. At each time point, the cells were trypsinized and counted with a hemocytometer.

*In vitro three-dimensional colony assays*

A 6-well plate was coated with 1.5 ml of 0.8% soft agar in a complete RPMI media; agar was solidified at 37°C with 5% $CO_2$ for 1 hour. A total of $3x10^5$ cells were placed into 1.5 ml of 0.4% soft agar in a complete RPMI media was plated on top of the solidified agar and then incubated at 37°C with 5% $CO_2$ for 24 hours. Complete RPMI media added on top of the agar and changed every 3 days. Each well was examined under microscope for colonies and were counted after 1, 2, and 3 weeks.

*Mouse colony and orthotopic transplant model*

All animal experiments were approved by the University of Iowa Institutional Animal Care and Use Committee (IACUC) and performed in accordance with IACUC guidelines. All mice used were female mice of BALB/c background purchased from Charles River (Wilmington, MA). Orthotopic transplants were performed as previously described.[269] Briefly, 6-week old mice were injected with 4T1-WT or 4T1-*Msh2* KO cells in 1:1 Matrigel/PBS were implanted into #4 mammary glands, bilaterally. Mice were treated with either 1x PBS as a vehicle control or 200 μg of anti-PD-1 antibody (Bio X Cell, West Lebanon, MA) every three days via intraperitoneal injection. The experiment was terminated when the largest tumors reached 2 cm in diameter.

*Immunohistochemistry*

Tumor sections were preserved in 10% neutral buffered formalin and embedded in paraffin. 5 μM thick sections were then deparaffinized with xylene and antigen was unmasked with Tris EDTA pH 9. Proteins was detected with anti-MSH2 (AbCam, Cambridge, MA), anti-FOXP3 (AbCam) and anti-CD3 (NeoMarkers, Freemont, CA) antibodies. Staining protocol was developed by the Comparative Pathology Core Facility in the Department of Pathology at the University of Iowa. Quantification of immune cells utilized blinded counting of at least 10 high-power fields of each tumor.

*Statistical analysis*

Two sample statistical testing principally utilized Welch's T test, allowing for unequal variance and sample size, unless otherwise indicated. Multiple comparisons across three or more groups utilized ANOVA with the Tukey Honest Significant Difference adjustment for multiple comparisons.

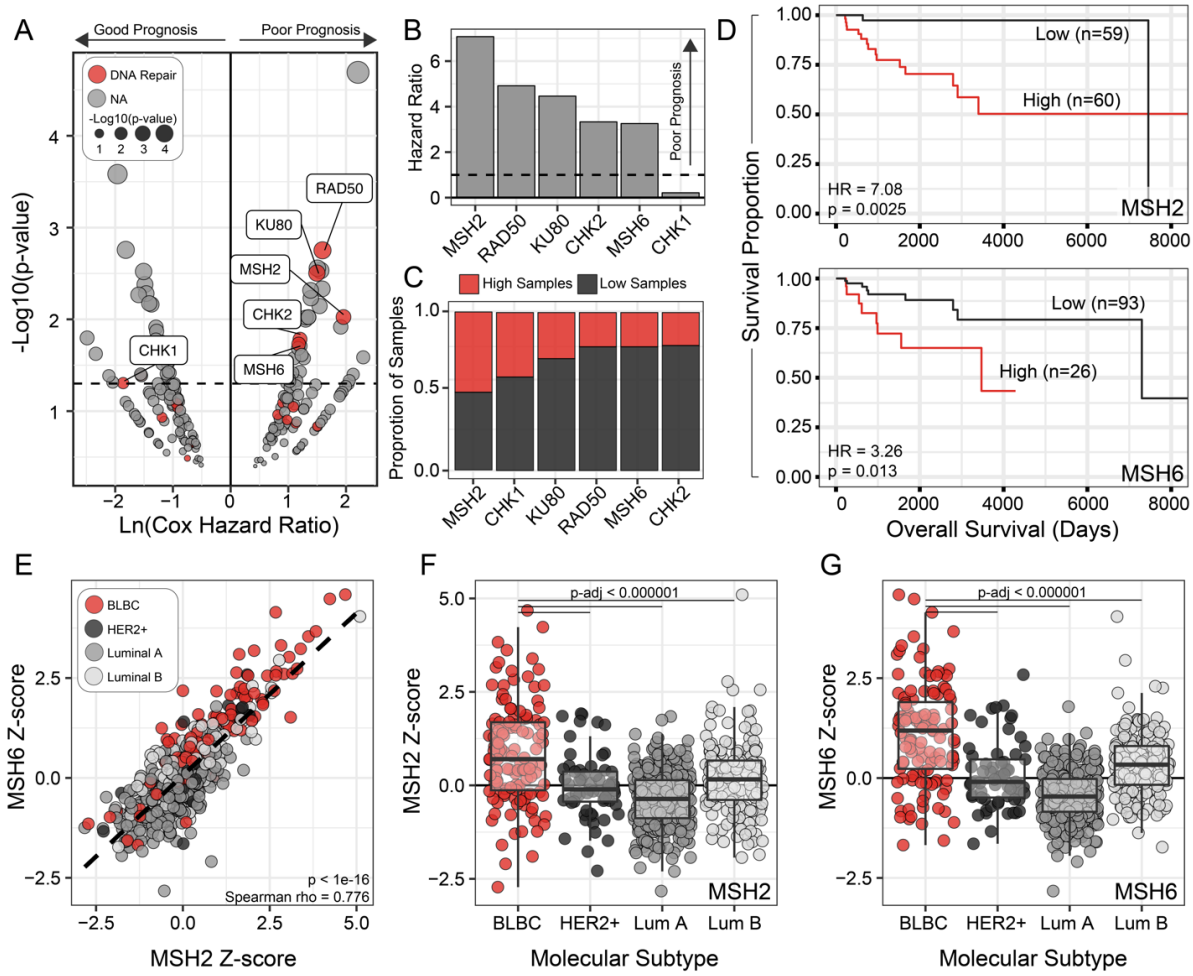


**Figure 6.** DNA mismatch repair proteins are elevated in BLBC and predict poor survival. **A.** Cox regression hazard ratio and p-values across all 224 proteins on the RPPA for 119 BLBC samples. DNA repair proteins are highlighted in red. **B.** Hazard ratio summary for the 6 RPPA-quantified DNA repair proteins with a p-value less than 0.05. Hazard ratio above one (dotted line) indicates poor survival for the higher protein level. **C.** Proportional cut points for the six significant RPPA proteins based on log-rank optimal p-value. **D.** Survival curve for MSH2 (upper panel) and MSH6 (lower panel) in 119 BLBC samples comparing high *versus* lo protein level. Significance based on log-rank test. **E.** MSH2-MSH6 protein correlations for breast cancer samples with molecular subtype designations (n=670, p < 1e-16). **F.** MSH2 RPPA protein quantification across molecular subtypes, BLBC (n=119), HER2+ (N=61), Luminal A (n=328), and Luminal B (n=162). p < 0.000001 for MSH2 in BLBC based on one-way ANOVA with Tukey HSD adjustment for multiple comparisons. **G.** MSH6 RPPA protein quantification across molecular subtypes. p < 0.000001 based on one-way ANOVA with Tukey HSD adjustment for multiple comparisons.

## Results

*MSH2 and MSH6 are elevated in BLBC and predict poor overall survival*

We previously developed analysis workflow to identify survival outcomes using RPPA quantification available in the Cancer Genome Atlas (TCGA).[262] We applied the system to examine prognostic markers for overall

survival in BLBC (**Figure 6A**), with a focus on DNA repair proteins available on the RPPA (highlighted in red). Of the 6 DNA repair proteins that had significant ($P < 0.05$) prognostic value, MSH2, RAD50, KU80, CHK2, and MSH6 predicted poor overall survival with Cox proportional hazard ratios greater than 1 (**Figure 6B**). The division of samples in high *versus* low was based on maximal standardized two-sample linear rank statistic to find an optimal cutpoint, providing a range of proportional comparisons (**Figure 6C**). Notably, MSH2 evenly split at 50% of samples (**Figure 6C**) and predicted the worse survival for any DNA repair protein (**Figure 6D**, upper panel, hazard ratio=7.08). MSH6, which forms the MutSα heterodimer with MSH2, had a different optimal cutpoint at 21.8% MSH6-high *versus* 78.2% MSH6-low, but still predicted poor overall survival (**Figure 6D**, lower panel, hazard ratio=3.26). Across breast cancer samples, MSH2 and MSH6 protein levels were directly correlated with one another (**Figure 6E**), supporting established literature on the need heterodimer formation for MSH6 stabilization.[270] Accompanying the survival predictions, we also found MSH2 (**Figure 6F**) and MSH6 (**Figure 6G**) are significantly elevated in BLBC compared to the other breast cancer molecular subtypes. We also performed Cox regressions in the other molecular subtypes, finding no significant predictive value for MSH2 or MSH6 in HER2+ or Luminal A breast cancers (**Figure 7A,B**). Interestingly, in Luminal B tumors, MSH2 protein level predicted improved overall survival (**Figure 7C**).

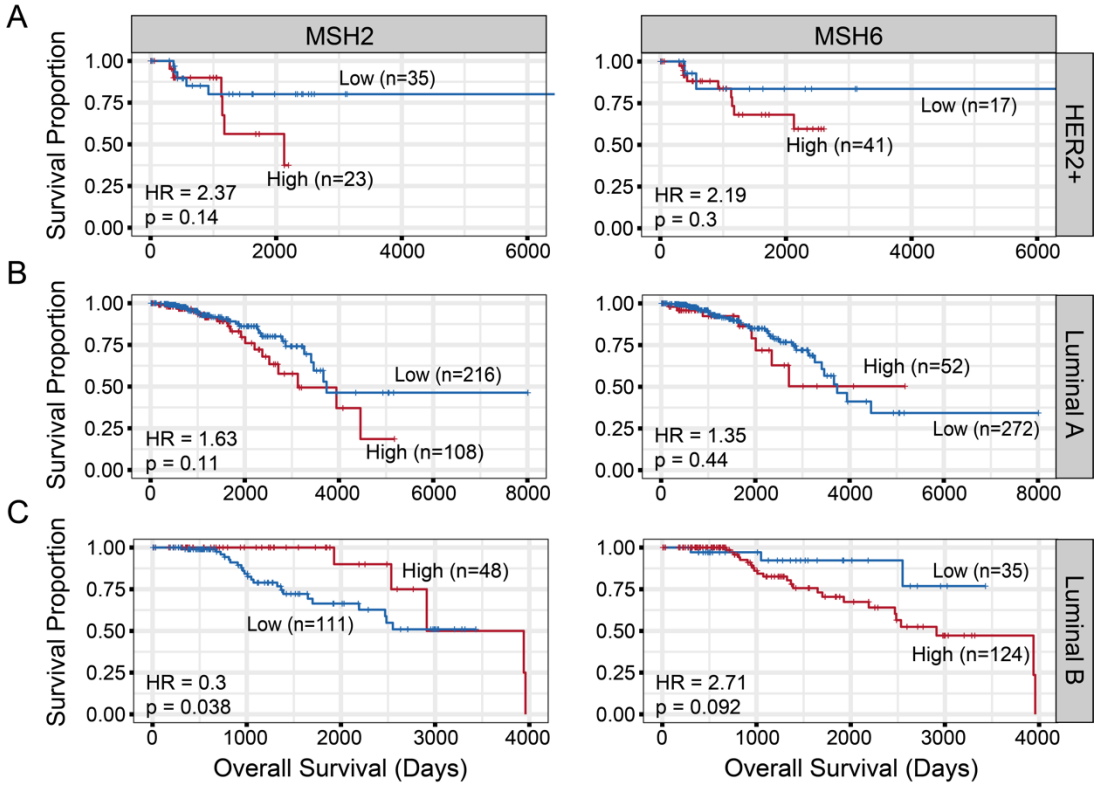


**Figure 7.** DNA mismatch repair prognosis in other molecular subtypes. **A**. Overall survival for MSH2 (left) and MSH6 (right) in HER2+ breast cancers (n=58). **B**. Overall survival for MSH2 (left) and MSH6 (right) in Luminal A breast cancers (n=324). **C**. Overall survival for MSH2 (left) and MSH6 (right) in Luminal B breast cancers (n=159). P-values calculated by log-rank testing with Cox proportional hazard regression modeling for hazard ratios (HR).

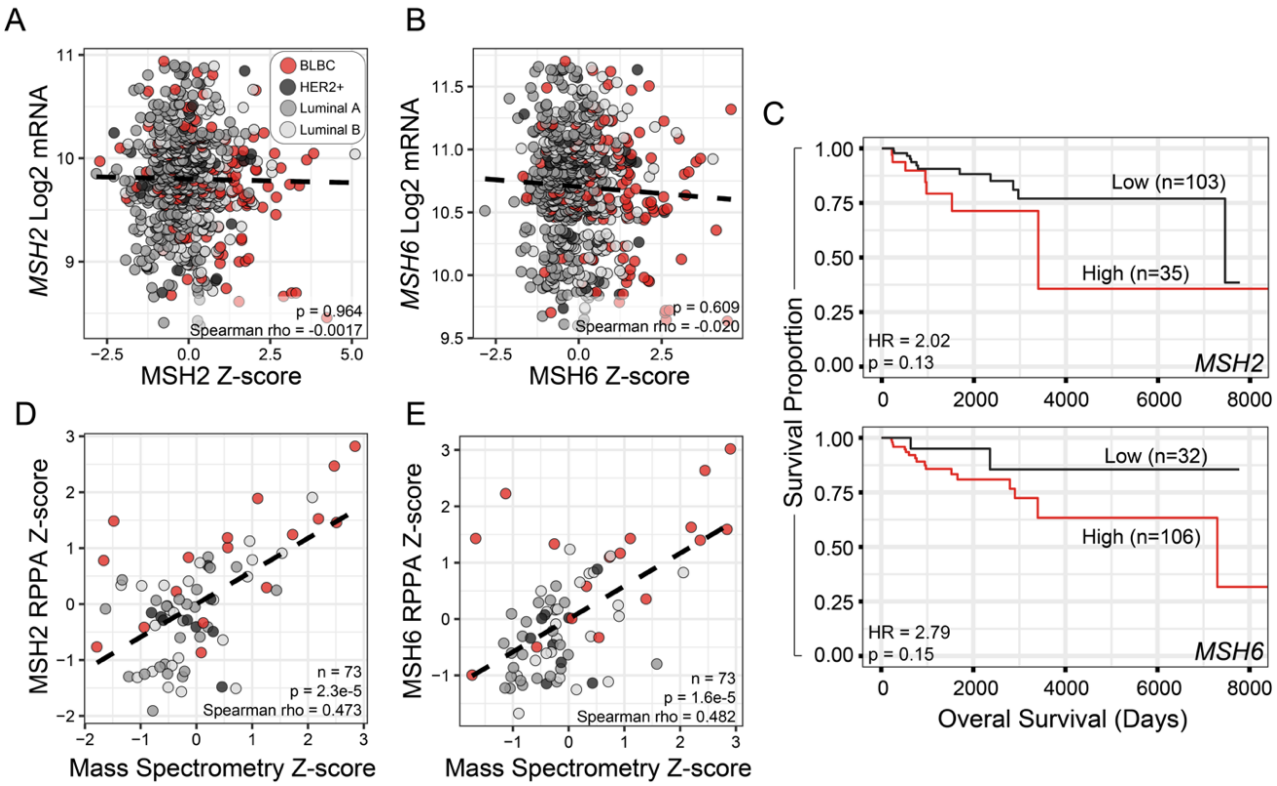


**Figure 8.** Mismatch repair mRNA and protein correlations in breast cancer. **A.** Spearman correlation of MSH2 RPPA Z-score *versus* *MSH2* mRNA level; BLBC (n=119), HER2+ (N=61), Luminal A (n=328), and Luminal B (n=162). **B.** Spearman correlation of MSH6 RPPA Z-score *versus* *MSH6* mRNA level. **C.** Survival curve for *MSH2* (top) and MSH6 (bottom) in 138 BLBC samples comparing high (red) to low (black) mRNA levels. **D.** Spearman correlation of MSH2 RPPA Z-score *versus* mass spectrometry-quantified MSH2 level; BLBC (n=17), HER2+ (N=8), Luminal A (n=27), and Luminal B (n=21). **E.** Spearman correlation of MSH6 RPPA Z-score *versus* mass spectrometry-quantified MSH6 level.

Due to the novelty of our finding of the poor prognostic indication of both MSH2 and MSH6 in BLBC, we further examine the dynamics of the MutSα constituents across the TCGA breast cancer samples. We first investigated if mRNA of *MSH2* (**Figure 8A**) or *MSH6* (**Figure 8B**) correlated with the respective RPPA probes, finding general poor linear correlations between protein and mRNA. This disparity between mRNA and protein level may have contributed to the lack of previous identification and could be a product of the post-translational regulation of DNA mismatch repair by ubiquitination.[271] As expected, the mRNA levels in BLBC (n=138) failed to find prognostic value for *MSH2* (**Figure 8C**, upper panel) or MSH6 (**Figure 8C**, lower panel) using the optimal cutpoint based on the log-rank p-value. In addition to mRNA levels, we also examined the correlation between RPPA probes and mass spectrometry values in a subset of the breast cancer TCGA samples (n=73) to better corroborate the use of RPPA-based protein levels.[272] In general, we found consistent correlations in both MSH2 (**Figure 8D**) and MSH6 (**Figure 8E**) with Spearman $\rho > 0.47$. With this confirmation of approximate protein levels using linear regressions of two separate measures of proteins, next we investigated differential gene regulation by DNA mismatch protein level.

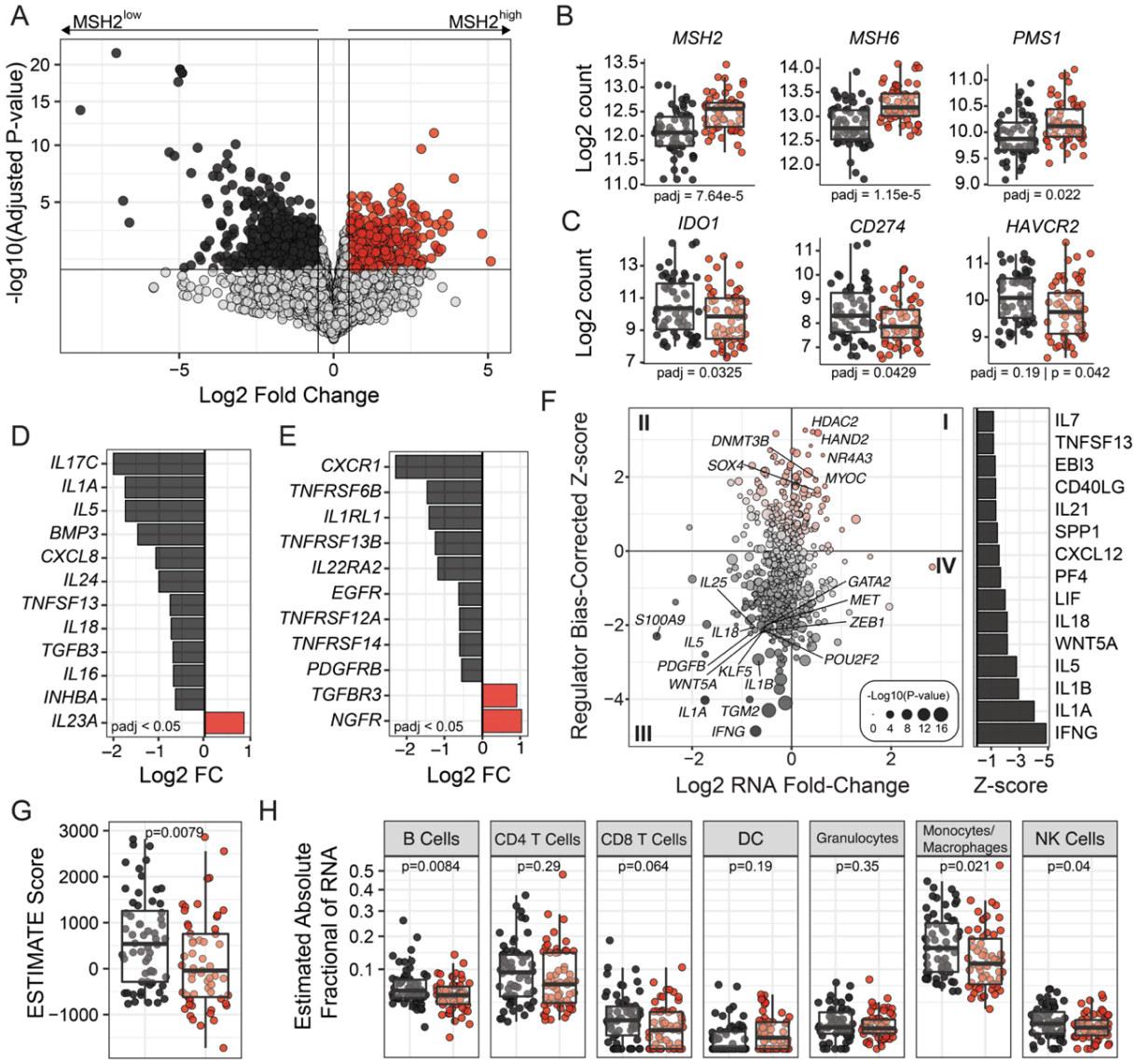


**Figure 9.** Differential gene expression results based on MSH2 protein in BLBC. **A.** Differential gene expression using HTSeq count data comparing MSH2-high (n=60) to MSH2-low (n=59) found 650 genes significantly increased and 1,444 significantly decreased. Significance was defined as an adjusted P-value < 0.05 and log2 fold-change ≥ |0.5|. **B.** Regularized log2-transformed count expression of DNA mismatch repair pathway constituents that demonstrated significantly differential between MSH2-high *versus* low. **C.** Regularized log2-transformed count expression of immune checkpoint targets. **D.** Log2-fold change of significant chemokines and cytokines with adjusted p-value < 0.05. **E.** Log2-fold change of significant chemokine and cytokine receptors with adjusted p-value < 0.05. **F.** IPA upstream regulators results based on differential gene expression. Quadrant I correspond to regulators with increased expression and predicted increase activity in MSH2-high samples, while Quadrant III corresponds to increased expression and predicted activity in MSH2-low samples. Bar graph summarizes the significant predicted activation of cytokines. **G.** Immune infiltration score based on the ESTIMATE algorithm by MSH2 mean protein value. **H.** Condensed CIBERSORT estimates of immune cell lineages based on mRNA values available for the BLBC samples.

*Increased MSH2 protein has significantly lower immune-associated genes*

As the common constituent of both MutSα (MSH2-MSH6) and MutSβ (MSH2-MSH3), we focused on MSH2 to examine the etiology of the difference in survival. To this end, we performed differential gene expression

compared MSH2-high (n=60, red) to MSH2-low (n=59, black) using estimated mRNA counts (**Figure 9A**). As expected, we found MSH2-high associated with increased RNA levels of *MSH2* and *MSH6*, in addition to increased *PMS1*, which participates in the repair downstream of the MutSα heterodimer (**Figure 9B**). Conversely, MSH2-low BLBC samples had increased levels of immune checkpoints, *IDO1*, *CD274*, and *HAVCR2* (**Figure 9C**). This increased levels of immune markers in MSH2-low samples was seen across chemokines/cytokines (**Figure 9D**) and chemokine/cytokine receptors (**Figure 9E**). Using the differential expression analysis, we also performed Ingenuity Pathway Analysis for upstream regulators of the transcriptional differences observed (**Figure 9F**). Quadrant I refers to upstream regulators that have increased mRNA expression and activation in MSH2-high BLBC samples, while quadrant III refers to regulators with increased expression and activation in MSH2-low samples. In MSH2-low BLBC samples, we observed a general increase in cytokine/chemokine signaling, notably in IL-1A/B and IFN-$\gamma$ (**Figure 9F**, quadrant III and adjacent bar graph). In addition we also used mRNA levels to estimate immune infiltration by MSH2 protein level using ESTIMATE and CIBERSORT algorithms.[266,267] Consistent with the differential gene expression analysis, we found increased estimated immune cell infiltration (**Figure 9G**), with lineage specific increases in B cells (P=0.0084), monocytes/macrophages (P=0.021), and NK cells (P=0.04), with $CD8^+$ T cells approaching significance (P =0.062) (**Figure 9H**).

*The knockout of Msh2 increases proliferation and migration of BLBC cells*

With preliminary indication of the modulation of anti-tumor immune responses dependent on DNA mismatch repair in BLBC, we developed two BLBC syngeneic cells lines for *in vivo* tumor modeling. The use of syngeneic murine cell line models allows for the evaluation of tumor growth in mice with intact immune function. Using CRISPR-Cas9 in a lentiviral vector and introducing a scrambled control or *Msh2*-specific guide RNA, we knocked out (KO) *Msh2* in 4T1 and Py8119 cells (**Figure 10A**). Before injecting the *Msh2*-WT or *Msh2*-KO cells into mice, we first characterized potential altered phenotypes *in vitro*. We first investigated changes in proliferation, finding *Msh2*-KO cells in both 4T1 and Py8119 consistently led to significant increases in proliferation (**Figure 10B**). In addition to proliferation, we observed a consistent increase in cell migration in *Msh2*-KO cells in both cell lines (**Figure 10C**,**D**). Both of these observations fit with the previous established role of DNA mismatch repair as a tumor suppressor. When we assayed invasion capabilities of *Msh2*-WT *versus* *Msh2*-KO cells, we found inconsistent increase in invasion of 4T1 *Msh2*-KO cells, while Py8119 *Msh2*-WT cells showed a trend towards increased invasion (**Figure 10E**,**F**). In contrast to the other *in vitro* findings, we found an decrease in growth capacity for *Msh2*-KO cells in three-dimensional growth assays using 4T1 (**Figure 10G**) and Py8119 (**Figure 10H**) cells. This decrease in three-dimensional growth is particularly intriguing as three-dimensional growth is thought to better recapitulates the tumor physiology and drug response.[273]

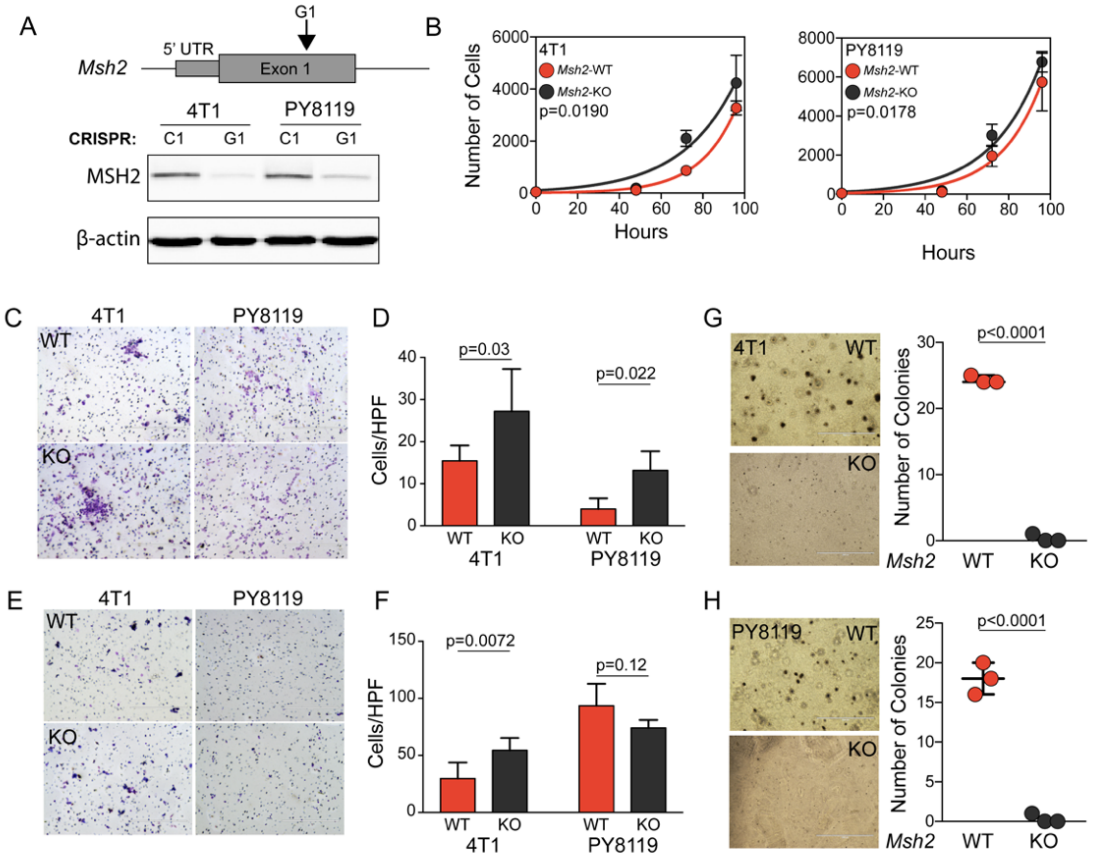


**Figure 10.** *Msh2*-KO in murine breast cancer alters *in vitro* growth. **A.** CRISPR-Cas9-mediated KO of *Msh2* in 4T1 and Py8119 murine breast cancer cell lines. **B.** Proliferation assay for $5x10^4$ 4T1 (left panel) and Py8119 (right panel) comparing *Msh2*-WT (red) to *Msh2*-KO (black) cells. Significance based on two-way ANOVA. **C**,**D.** Representative images (**C**) and quantifications (**D**) of transwell migration assays for *Msh2*-WT (red) to *Msh2*-KO (black) cells. **E**,**F.** Representative images (**E**) and quantifications (**F**) of transwell invasion assays for *Msh2*-WT (red) to *Msh2*-KO (black) cells. **G**,**H.** Representative images and quantification of three-dimensional growth assays for 4T1 (**G**) and Py8119 (**H**) for *Msh2*-WT (red) to *Msh2*-KO (black) cells. Significance based on Student's T test. *In vitro* assays were performed by Dr. Edward Cho.

*Msh2-KO potentiates immune checkpoint blockade in vivo*

Previous research has found remarkable response rates of 25-80% for immune checkpoint blockade in Lynch Syndrome patients with defective DNA mismatch repair.[257–259] In order to asses if the genetic ablation of MSH2 could function as an adjuvant for checkpoint blockade in BLBC, we injected $10^5$ 4T1 *Msh2*-WT or *Msh2*-KO cells into the mammary glands of 6-week-old BALB/c mice (**Figure 11A**). Upon injection, the mice received intraperitoneal injections of vehicle control or 200 μg of anti-PD-1 antibody every three days for the first two weeks. We found the *Msh2*-KO cells had a significant reduction in tumor growth compared to *Msh2*-WT cells (**Figure 11B**). The anti-PD-1 therapy reduced tumor growth in both *Msh2*-WT and *Msh2*-KO cells, but was additive to the reduction of tumor growth for *Msh2*-KO alone (**Figure 11B**). Similar to tumor growth, both the anti-PD-1 treatment and the KO *Msh2* significantly increased survival, with the *Msh2*-KO receiving anti-PD-1 having a median survival time of 58 days post injection compared to the 48 days of both the *Msh2*-KO + vehicle or *Msh2*-WT + anti-PD-1 groups (**Figure 11C**).

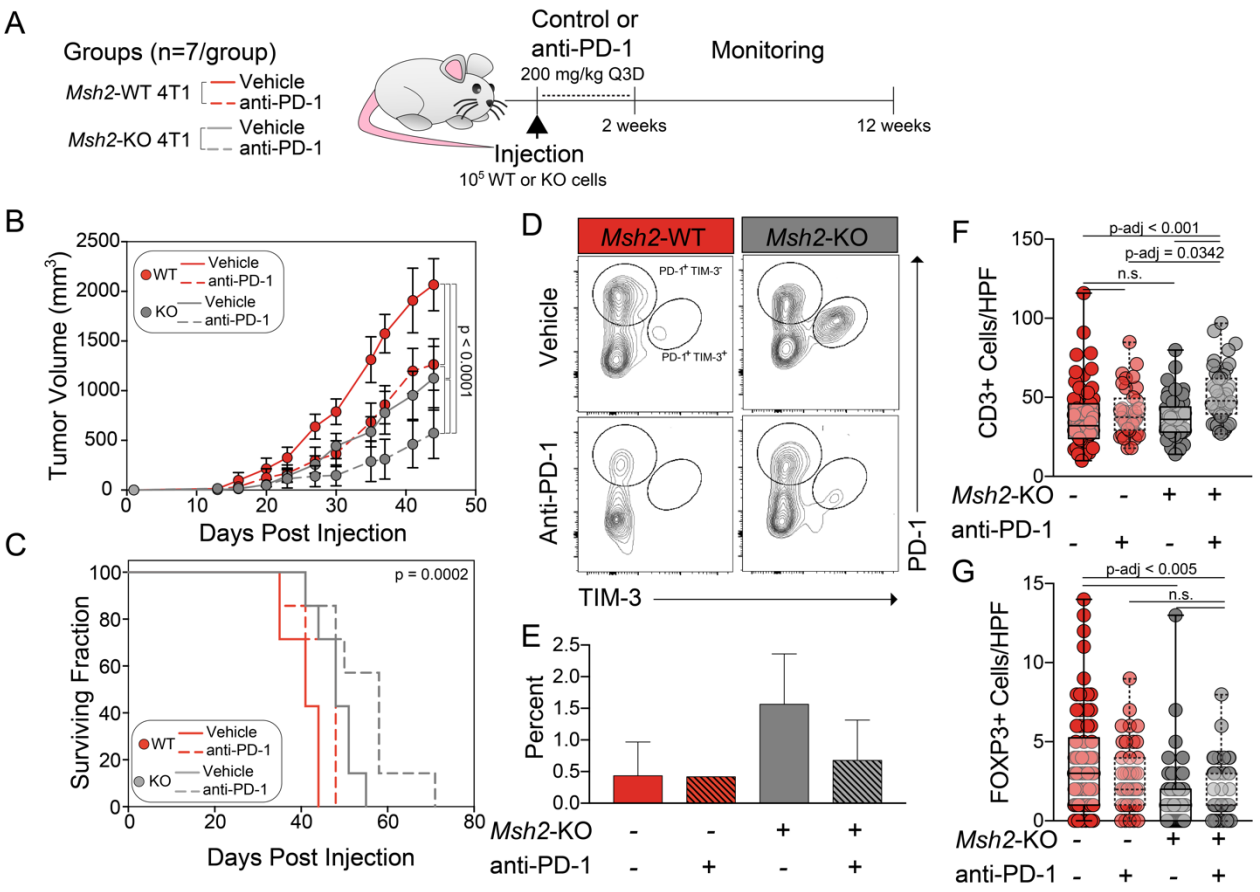


**Figure 11.** *Msh2*-KO potentiates immune checkpoint blockade *in vivo*. **A.** Schematic of *in vivo* tumor study using 4T1 *Msh2*-WT *versus* *Msh2*-KO with or sans anti-PD-1 therapy. **B.** Tumor growth curve of *Msh2*-WT (red) *versus* *Msh2*-KO (grey) with anti-PD-1 therapy (dashed line) or vehicle (solid line). Significance based on two-way ANOVA with individual comparisons corrected for multiple hypothesis testing using the Tukey HSD method. **C.** Survival curve for the four arms of the experiments. Significance based on log-rank test. **D.** Representative flow cytometry plots of tumor-infiltrating CD4+ T cells, stained for PD-1 (y-axis) and TIM-3 (x-axis). **E.** Percent of tumor-infiltrating PD-1+ TIM-3+ CD4+ T Cells compared to total tumor-infiltrating CD4+ T Cells by group. Flow cytometric measurements were performed by Dr. Gaurav Pandey. **F.** Quantification of CD3+ cells by IHC staining across the four groups using at least 10 high-powered fields (HPF) per sample. Significance based on one-way ANOVA using the Sidak method for multiple comparison correction. **G.** Quantification of FoxP3+ cells by IHC staining across the four groups using at least 10 high-powered fields (HPF) per sample. Significance based on one-way ANOVA using the Sidak method for multiple comparison correction.

Immune checkpoint therapy, like anti-PD-1 antibodies, are thought to work, in part, through the prevention of T cell exhaustion.[259] To assess the modulation of infiltrating T cell phenotypes, we isolated CD3+ T cells populations from tumors using flow cytometry. Although we did not find a clear difference in CD8+ T cells (data not shown), we did find a decrease in PD-1+ TIM-3+ CD4+ T cells with anti-PD-1 therapy in both *Msh2*-WT and *Msh2*-KO groups (**Figure 11D**,**E**). Both PD-1+ and TIM-3+ are upregulated upon T cell activation and are used to identify both activated and exhausted T cells.[95] We observed a substantial increase in PD-1+ TIM-3+ CD4+ T cells in the *Msh2*-KO + vehicle group that was ameliorated by the addition of anti-PD-1 therapy (**Figure 11D**,**E**). Using IHC staining using a pan-T cell marker, CD3, we also observed a significant increase in the average number of T cells per high-power field (HPF) in *Msh2*-KO + anti-PD-1 compared to all other

groups (**Figure 11F**). In addition to CD3, we also examined FoxP3$^+$ cells per HPF in all four groups, finding a reduction in the *Msh2*-KO groups compared to the untreated *Msh2*-WT control (**Figure 11G**). A marker and master regulator of regulatory T cells (Tregs), a reduction in tumor-infiltrating Tregs has been shown to be a good prognostic indicator in breast cancer.[274]

**Discussion**

BLBC is a particularly aggressive form of breast cancer with higher rates of metastasis and relapse compared to other subtypes. Despite recent progress in targeted therapies for other breast cancers, BLBC/TNBC is still treated principally with surgery, radiation, and broad cytotoxic chemotherapies.[275–277] In an effort to develop durable therapies for BLBC, investigation has led to the immunogenic nature of the subtype. Previous reports have found increased tumor-infiltrating lymphocytes in TNBC compared to other breast cancer subtypes.[278] Moreover the increased infiltration of lymphocytes in TNBC has been correlated with higher likelihood of benefit from adjuvant and neoadjuvant chemotherapy, with lower risk of disease relapse.[279–283] Despite these observations that suggest an immunogenic character in TNBC, initial results of anti-PD-1 therapy across all TNBC patients had response rates ranging from 5-8.8%.[284,285]

Here we identified constituents of the MutSα complex, MSH2 and MSH6, at the protein level were poor prognostic indicators in BLBC (**Figure 6D**), but not other breast cancer subtypes (**Figure 7**). Within BLBC samples, MSH2 protein was associated with a decrease immune cell signatures (**Figure 9G**,**H**) and decreased expression of immune checkpoints (**Figure 9C**). Conversely, a hallmark of tumors with defective mismatch DNA repair is dense peri- and intratumoral infiltration of T cells.[209] and responsiveness to immune checkpoint blockade.[257–259] T cells infiltrate in tumors with dysfunctional mismatch DNA repair are primed for cytotoxicity and have reduced suppression by Tregs.[286–288] Thus, this increased expression of the MutSα complex in BLBC compared to other subtypes (**Figure 6F**,**G**) may serve to offset the high levels of genomic instability and reduce anti-tumor immune resposes.[289]

In terms of the latter inhibition of anti-tumor response, we demonstrate the genetic KO of *Msh2* in syngeneic tumor models led to a reduction in tumor growth and increased survival (**Figure 11B**,**C**). Indicative of the immunomodulation phenotype of the *Msh2*-KO model was the increase in double-positive PD-1$^+$ TIM-3$^+$ CD4$^+$ T cells (**Figure 11D**) and a reduction in Tregs numbers by FoxP3 IHC (**Figure 11G**). PD-1 and TIM-3 serve as negative immune checkpoints and are upregulated upon T-cell-receptor activation, serving as markers of both activation and exhaustion.[290] The blockade of the interaction of PD-1 to PD-L1 or PD-L2 is thought to function by preventing T cell exhaustion and maintain effector T cell function.[72,259] Beyond the negative regulation of T cell activation, TIM-3 also appears to direct immune responses away from the anti-tumor Th1 response,[291] and is currently being investigated in early clinical trials with anti-PD-1 therapies to improve overall

response rates to immune checkpoint blockade. Interestingly, the addition of anti-PD-1 therapy diminished the double-positive PD-1$^+$ TIM-3$^+$ CD4$^+$ T cells in the *Msh2*-KO model (**Figure 11D**) and led to additive effects on reducing tumor growth and increasing the length of overall survival (**Figure 11B**,**C**).

A number of clinical trials are currently underway in order to identify combinatorial approaches to potentiate or promote immune checkpoint blockade across a solid tumors.[259] Careful consideration of combination strategies is necessary for broadening the patients that will benefit from immune checkpoint blockade. PARP inhibitors of single-strand DNA repair, have shown to improve the response rate to checkpoint inhibitors.[292] Similarly, the use of platinum agents with immune checkpoint blockade in lung cancers improve clinical outcomes for patients.[293,294] Our preliminary work has demonstrated the viability of the removal of mismatch DNA repair is sufficient to act as an immune adjuvant for anti-PD-1 therapy in BLBC. With low levels of overall response to anti-PD-1 therapy in BLBC, further work on the underlying mechanism of action and potential inhibitors of the MutSα complex is warranted.

## CHAPTER 3
## HETEROGENEITY AND FUNCTIONAL IMPLICATIONS OF CANCER-ASSOCIATED REGULATORY T CELLS

### Rationale

Tregs are a population of T cells that exert suppressive effects on a variety immune cells including $CD8^+$ T cells, $CD4^+$ T cells, natural killer cells, and dendritic cells.[295] Tregs play an indispensable role in maintaining normal immune homeostasis and peripheral tolerance. However, their suppressive effect in the tumor microenvironment is detrimental[295] prompting Treg-targeted immunotherapy.[296] Ipilimumab, the first immune checkpoint blocker approved by the FDA, is thought to work through inhibiting Treg function by abolishing CTLA-4-mediated suppressive signaling or preferentially depleted Tregs in the tumor microenvironment.[297–299] Tregs are identified by the expression of transcription factor forkhead box P3 (*FOXP3*), which acts as a master regulator for Treg development and the suppressive functions of Tregs.[300,301] However, *FOXP3* alone fails to predict patient survival and its prognostic value was highly tumor-type- and stage-dependent across different tumors.[274,302,303] Moreover, recent studies unveiled functional heterogeneity of $FOXP3^+$ Tregs in peripheral blood[304] as well as in different tumor types including colorectal cancer[305] and glioma.[306] General disruption of Treg function negatively impacts immune homeostasis, as seen in the autoimmune complications associated with CTLA-4 blockade and underscores the need for refined markers and targets for tumor-infiltrating Tregs.

Recent mRNA sequencing of blood-derived, normal-tissue-resident, and tumor-infiltrating Tregs from different tumor types in two independent studies revealed sizable transcriptional overlap between normal-tissue and tumor-infiltrating Tregs.[307,308] Nevertheless, tumor-infiltrating Tregs showed differential upregulation of appreciable number of genes, including, chemokine receptor 8 (*CCR8*), cytotoxic T-lymphocyte associated protein 4 (*CTLA4*), lymphocyte activation gene 3 (*LAG3*), and T-cell membrane protein 3 (TIM-3, encoded by *HAVCR2*), layilin (*LAYN*), and MAGE family member H1 (*MAGEH1*).[307,308] Relying on sequencing of pooled tumor-infiltrating Treg, these studies fell short in untangling the heterogeneity of tumor-infiltrating Tregs reported earlier at the functional level. In a more recent study, approximately 5,000 flow-sorted T cells from tumor, normal tissue, junctional tissue and blood of patients with hepatocellular carcinoma (HCC) were analyzed by single-cell RNA sequencing.[309] In that study, initial analysis uncovered $CD4^+$ and $CD8^+$ T cell heterogeneity at the transcriptional level suggesting functional heterogeneity of tumor-infiltrating T cells.

In order to investigate heterogeneity and dynamics of cancer-associated Tregs, we perform single-cell sequencing on peripheral and tumor-infiltrating immune cells in three renal clear cell carcinoma (ccRCC) patients (**Figure 12**). We chose ccRCC tumors based on the responsive of these tumors to immune checkpoint blockade despite low mutational loads, which implies the importance of the tumor microenvironment.[180] Pairing our novel dataset with the single-cell sequencing of the immune cells of the HCC dataset, we identified

*CD177* associated with a suppressive subset of tumor-infiltrating Tregs. CD177 (also known as NB1, HNA-2a, or PRV1) is a glycosylphosphatidylinositol-linked cell surface protein that is expressed heterogeneously by neutrophils in 90% of humans[310] and has been identified as a useful biomarker for myeloproliferative diseases.[311] We recently generated a mouse model with knockout (KO) of *Cd177* and demonstrated that CD177 plays a role in neutrophil viability.[312] Limited literature indicates a correlation between loss of *CD177* expression and poor prognosis in colorectal and gastric cancer.[313,314] Recently, a single report by the Rudensky group identified CD177$^{+}$ tumor-infiltrating Tregs in breast cancer with no further elaboration.[308] Using machine-learning techniques we developed a gene signature for the suppressive Treg subset that predicts poor overall survival across a number of cancer types. Further understanding of the development and regulation of this suppressive subset of tumor-initiating Tregs may be crucial in the development of cancer immunotherapies with minimal auto-immune side effects.

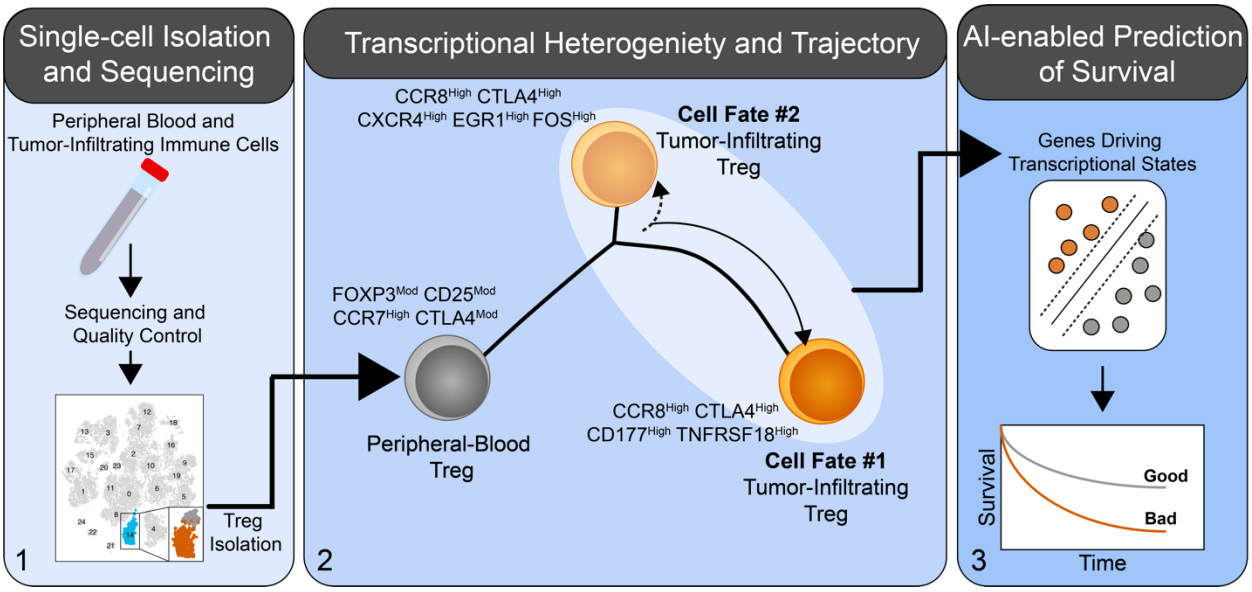


**Figure 12.** Schematic summary of the Treg heterogeneity analysis and predictions. **1.** Single-cell RNA sequencing was performed on peripheral blood (n=13,433) and tumor-infiltrating immune cells (n=12,239) from 3 different ccRCC patients. **2.** From these cells, Tregs were isolated and analyzed for markers of tumor-infiltrating Tregs and analysis of heterogeneity. **3.** Signatures developed from the heterogeneity outperformed previous approaches to predicting survival based on mRNA levels.

## Materials and methods

*Patient recruitment*

The current study was approved by the University of Iowa Institutional Review Board (IRB) and conducted under the Declaration of Helsinki Principles. Deidentified ccRCC were recruited by Dr. Yousef Zakharia in the Department of Internal Medicine at the University of Iowa. Glioma, melanoma, pancreatic adenocarcinoma, and spleen samples were provided by Dr. Yuwen Zhu from University of Colorado. These tissues were also

collected from patients undergoing surgical resection after informed consent and were supplied as deidentified samples to Dr. Zhu laboratory with IRB approval.

*Single-cell RNA sequencing*

Viable (Hoechst$^-$), immune (CD45$^+$) single-cell suspensions generated from three ccRCC tumor samples and peripheral blood were sorted on a FACS ARIA sorter (BD Biosciences) for lymphoid and myeloid lineages. The cells were sorted into ice-cold 1x PBS + 0.04% non-acetylated bovine-serum albumin (New England BioLabs, Ipswitch, MA). Sorted cells were counted and viability as assessed using the MoxiGoII counter (Orflo Technologies, Ketchum, ID). Single cells were re-suspended at 1000 cells/μl with a viability > 90%. Sequencing was performed using the Chromium (10x Genomics, Pleasanton, CA) and Illumina (San Diego, CA) sequencing platforms. Amplified cDNA was used to construct both 5' expression and VDJ sequencing libraries. Pooled libraries were run on sperate lanes of a 150 based-paired, paired-end, flow cell using the Illumina HiSeq 4000. Basecalls were converted into FASTQs using the Illumina bcl2fastq software by the University of Iowa Genomics Division. FASTQ files were aligned to human genome (GRCh38) using the CellRanger v2.2 pipeline as described by manufacturer.[315] Individual cells were adjusted for total expression and percentage of mitochondrial gene expression. After processing, clustering was performed using the Seurat package (v2.3.4), correcting for patient variability using the canonical correlation process.[316,317] Dimensional reduction to form the tSNE plot utilized the top 20 calculated dimensions and a resolution of 1.2. Differential gene expression analysis was performed using the Wilcoxon rank sum test for significance comparing tumor-infiltrating *versus* peripheral-blood Tregs. Cell trajectory manifolds and pseudo-time estimates utilized the Monocle R package (v2.8.0) and the reverse graph embedding machine learning algorithm.[318] The same processing and quality control procedures were applied to count-level data from GSE98638.[309]

*Bulk RNA sequencing of Tregs*

Raw expression data for GSE89225[308] and PRJEB11844[307] were downloaded from the NCBI Sequence Read Archive and the European Nucleotide Archive, respectively. SRA files were converted to FASTQ files using the SRA toolkit. Samples were aligned with the kallisto pseudoalignment protocol and GRCh38 build for the human genome to produce estimated counts.[319] Treg expression values were processed using the Sleuth R Package (v0.30.0).[320]

*Cell lines and cell culture*

PY8119 cells were derived from a primary *MMTV-PyMT* tumor in the C57BL/6 background as described previously[321] and were maintained in F12 media supplemented with 10% fetal bovine serum (FBS), 10 ng ml$^{-1}$ epithelial growth factor, 2 μg ml$^{-1}$ hydrocortisone and 5 μg ml$^{-1}$ insulin. When needed, cells were detached using 2.5% trypsin at P18 and resuspended in 1:1 phosphate buffer saline: Matrigel (Corning, Corning,

NY) mixture and injected in mammary fatpad at 100 µl volume per inoculation. MC38 colorectal cells were maintained in the same complete F12 media, detached using 2.5% trypsin, and resuspended in phosphate buffer saline and injected into mammary fatpad at 100 µl volume per inoculation.

*Animals*

All animals were maintained under specific pathogen-free conditions according to the IACUC guidelines. We purchased mouse sperm carrying the *Cd177* deletion allele from the UC Davis KOMP Repository and sent it to Jackson Laboratories for *in vitro* fertilization. Seven mice were obtained with 3 carrying a heterozygous deletion of *Cd177*. Mice were backcrossed to C57BL/6 for 6 generations. Genotyping of the *Cd177*-KO mice was performed through standard PCR procedures. More information on the development and characteristics of the *Cd177*-KO mouse is available in our previous publication.[312]

*Tissue collection and T cell suppression assay*

Fresh human breast cancer and renal cancer tumor samples were collected from patients undergoing surgical resection after informed consent, detailed above. Breast tumor samples were supplied de-identified by the Tissue Procurement Core at the University of Iowa Hospitals and Clinics and were histologically characterized by the Department of Pathology at the University of Iowa according to immunohistochemical ER, PR and HER2 biomarker staining. De-identified blood samples from healthy donors were provided by the blood bank at the University of Iowa Hospitals and Clinics. For renal cancer patients, de-identified matching blood samples were provided by the Genito-Urologic Molecular Epidemiology Resource at the Holden Comprehensive Cancer Center, University of Iowa Hospitals and Clinics. Human peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using Ficoll Plaque (GE Healthcare Biosciences, Pittsburgh, PA) density gradient centrifugation or SepMate Tubes (StemCell Technologies).

Fresh tissues were directly distributed to the research laboratories after surgery, followed by enzymatic digestion and physical dissociation using gentleMACS (Miltenyi, Bergisch Gladbach, Germany) as per manufacturer's instruction. Cell suspensions were filtered through 100-µm cell strainer, magnetically enriched using anti-CD45 positive selection (Miltenyi) for tumor-infiltrating leukocytes, and immediately frozen using FBS with 5% DMSO. Breast cancer samples were used fresh or frozen without CD45 enrichment.

For *in vitro* suppression assay, naive $CD4^+$ T cells were isolated from PBMC using the human naïve $CD4^+$ T cell isolation kit II (Miltenyi, 130-094-131). Tumor-infiltrating Tregs were labeled with anti-CD45, anti-CD4, anti-CD25, and anti-CD177 antibodies and flow-sorted and combined from 3 human breast cancer specimens to get enough number of cells, $CD45^+CD4^+CD25^+$ T cells were used as total tumor-infiltrating Tregs, in addition to further separation of $CD177^+$ or $CD177^-$ tumor-infiltrating Tregs. Naïve $CD4^+$ T cells

and tumor-infiltrating Tregs were co-cultured at the indicated ratios (2:1, 4:1, or 8:1) on 96-well flat-bottom plates in RPMI1640 supplemented with 10% fetal bovine serum (FBS), 10 mM HEPES, $2x10^{-5}$ M 2-mercaptoethanol for 96 hr. Naïve $CD4^+$ T cells were labeled with CFSE and stimulated by anti-CD3/28 dynabeads (Thermo Fisher Scientific) and 10ng/ml IL-2 (R & D), following with flow cytometry to determine T cell proliferation indicated by CFSE dilution

*Tumor-infiltrating lymphocyte isolation*

For tissues used for flow cytometry staining, freshly resected tumor specimens were manually diced into 1mm pieces, followed by digestion with collagenase IV (Sigma) 1 mg/ml solution in 2% FBS in RPMI media at 1 ml per 0.1 g tissue for 1.5 hours at 37°C in a shaker incubator at 100 rpm. Tumor homogenates were passed through a 100-μm cell strainer and stained immediately or frozen in 5% DMSO in FBS for future use.

*Flow cytometry and antibody staining*

Multicolor phenotypic panels were established using different combinations of fluorescently tagged anti-CD45 (H130), CD3 (HIT3a), CD4 (OKT4), CD25 (M-A251), CD27 (O323), CD127 (A019D5), CCR8 (433H), PD-1 (EH12.2H7), CTLA-4 (BN13), FOXP3 (206D), and CD177 (MEM-166). Cells were stained using standard immunofluorescent staining protocol and run on flow cytometry either as live cells or fixed in 4% paraformaldehyde. Antibodies were purchased from Biolegend, Molecular probes, BD Biosciences and eBioscience. Flow cytometric data was acquired on a 4-laser, 19-parameter BD LSR II and data were analyzed using FlowJo software (TreeStar, Ashland, OR). For experiments using frozen samples, cells were thawed and suspended in RPMI supplanted with 10% FBS and incubated for 1.5 h at 37°C, 5% $CO_2$ prior to staining. Tregs were identified as $CD3^+CD4^+FOXP3^+$ or $CD3^+CD4^+CD25^+CD127^{low}$.[304]

*Signature development and evaluation*

Feature selection was performed for three sets of genes: 1) 143 differentially-expressed genes between tumor-infiltrating Tregs and peripheral-blood Tregs shared between ccRCC and HCC, 2) 86 genes differentially expressed in the second cell fate, and 3) 222 genes differentially expressed in the first cell fate. For each set of feature-selected genes, support vector machines (SVMs) were trained to predict overall survival using the e1071 (v1.7-0) R package. The 538 patients in the KIRC dataset were randomly divided into training (N=266) and testing (N=267) cohorts. SVMs were trained to discriminate overall survival status with linear kernels, and the cost-parameter was selected via cross-validation. Kaplan-Meyer curves were constructed with the survival (v2.42.6) and survminer (v0.4.3) R packages. The cox proportion hazards regression model within the survival package was used to compute the hazard ratios between good-outcome and poor-outcome prediction groups.

## Results

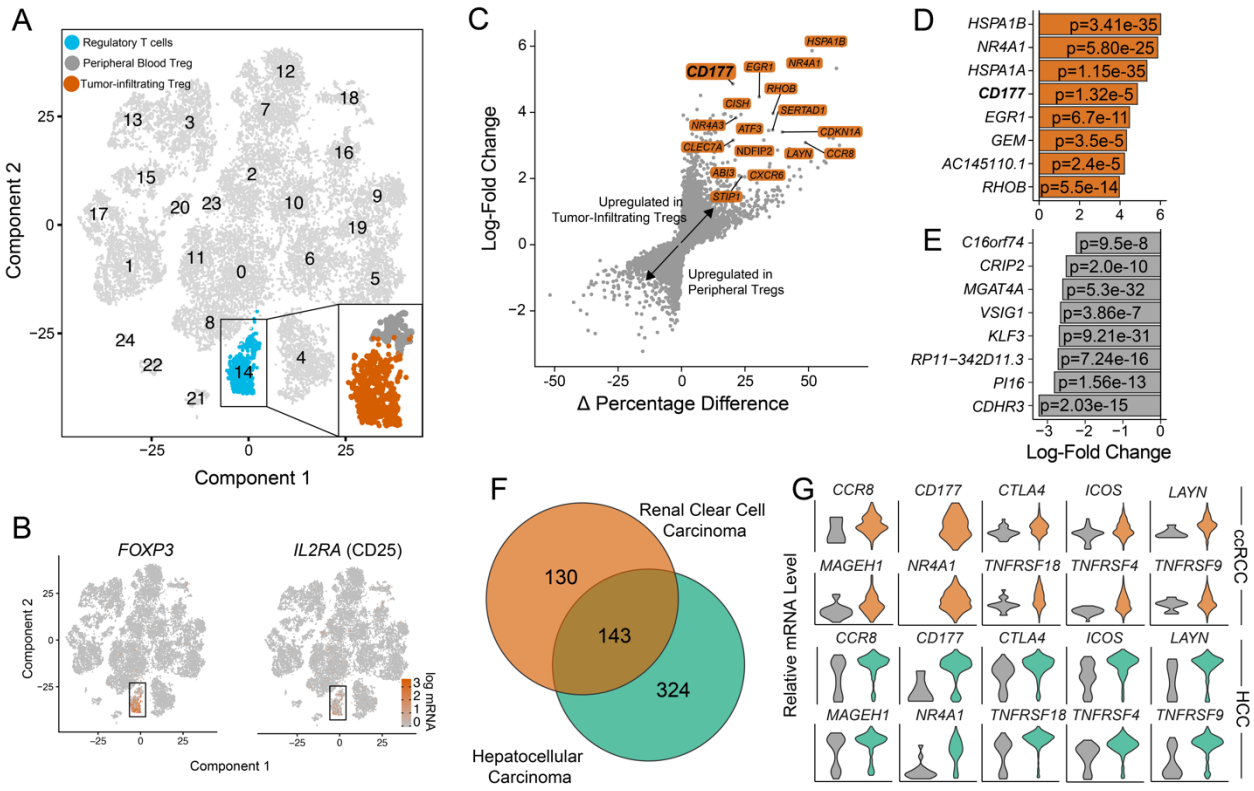


**Figure 13.** Single-cell sequencing and isolation of Tregs. **A.** tSNE projection of immune cells from three ccRCC patients with normal peripheral blood cells (n=13433) and tumor-infiltrating cells (n=12,239). Treg population (blue) was isolated and separated as tumor-infiltrating (orange) *versus* peripheral-blood Tregs (grey). **B.** tSNE projection with highlighted expression of Treg markers, *FOXP3* and *IL2RA* (CD25). **C.** Differential gene expression analysis using the log2-fold change expression *versus* the difference in the percent of cell expressing the gene comparing tumor-infiltrating minus peripheral-blood Tregs (Δ percentage difference) Genes labeled have log2-fold change > 1, Δ percentage difference > 20% and adjusted P-value from Wilcoxon rank sum test < 0.05. **D.** Top eight upregulated genes by log2-fold change in tumor-infiltrating Tregs with adjusted P-value < 0.05. **E.** Top eight downregulated genes by log2-fold change in tumor-infiltrating Tregs with adjusted P-value < 0.05. **F.** Comparison of differential genes in tumor-infiltrating Tregs in ccRCC (orange) and HCC (green) compared to peripheral-blood Tregs. Significant genes were defined as log2-fold change > 1 or < -1 with adjusted P-values < 0.05. **G.** Relative mRNA levels of Treg markers that were consistently increased in both ccRCC- and HCC-infiltrating Tregs, Log2-fold change > 1 and adjusted P-values < 0.05.

### *Single-Cell RNA sequencing in Tregs from ccRCC patients*

A total of 25,672 immune cells across the three ccRCC were isolated and passed quality control, with 13,433 peripheral-blood and 12,239 tumor-infiltrating cells. We isolated Tregs using the expression of the markers *FOXP3* and CD25 (*IL2RA*) in addition to the cluster identity from the t-distributed stochastic neighbor embedding (tSNE) method (**Figure 13A**,**B**). In total, we recovered single-cell-level expression profiles of 160 peripheral-blood Tregs and 574 tumor-infiltrating Tregs. In order to identify differentially-expressed genes and potential markers of tumor-infiltrating Tregs, we performed the Wilcoxon rank sum test comparing tumor-infiltrating *versus* peripheral-blood Tregs. In addition to using log2-fold change, we also examined the difference in the percentage of cells that express the specific gene in each category (Δ Percentage Difference), allowing for more specific evaluation of tumor-infiltrating Treg markers (**Figure 13C**). These tumor-infiltrating-specific

genes with a log2-fold change ≥ 1, adjusted P < 0.05 and a Δ Percentage Difference > 20% are labeled in **Figure 13C**. The only two genes with 0% expression in peripheral-blood Tregs were *NR4A1* (tumor-infiltrating percentage = 51.4%) and *CD177* (tumor-infiltrating percentage = 20.2%) (**Figure 13C**). A summary of the top eight genes by log2-fold change which are upregulated (**Figure 13D**) and downregulated (**Figure 13E**) in tumor-infiltrating Tregs are shown with the corresponding adjusted P-value.

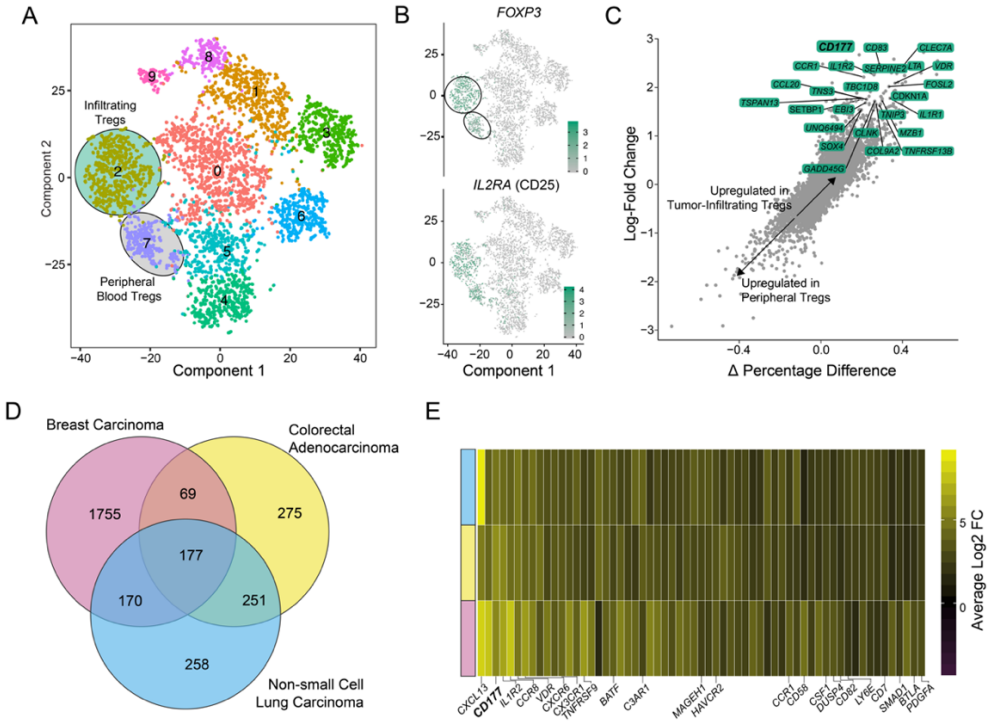


**Figure 14.** Processing of secondary tumor-infiltrating Treg data sets. **A.** tSNE projection of immune cells from five HCC patients with normal peripheral blood cells and tumor-infiltrating T cells. **B.** tSNE projection with highlighted expression of Treg markers, *FOXP3* and *IL2RA* (CD25). **C.** Differential gene expression analysis using the log2-fold change expression *versus* the difference in the percent of cells expressing the gene comparing tumor-infiltrating minus peripheral-blood Tregs (Δ Percentage Difference) Genes labeled have log2-fold change > 1, Δ Percentage Difference > 15% and adjusted P-value from Wilcoxon rank sum test < 0.05. **D.** Venn diagram of the overlap in differentially-expressed tumor-infiltrating Tregs compared to peripheral-blood Tregs in breast carcinoma (GSE89225), colorectal adenocarcinoma (PRJEB11844), and non-small cell lung squamous carcinoma (PRJEB11844). **E.** Heatmap differential genes shared between the three datasets with immune-related genes labeled. Genes are displayed in log2-fold change (FC) comparing tumor-infiltrating Tregs *versus* peripheral-blood Tregs with FDR q-values <0.05.

In order to compare our differential gene expression results for tumor-infiltrating Tregs, we used a recently-published single-cell profiling of T cells in HCC.[309] This data set is comprised of flow-sorted CD4+, CD8+, and Tregs from 5 HCC patients with cells isolated from peripheral blood, normal liver parenchyma, transitional zone near the tumor, and tumor-infiltrating T cells. A total of 4,922 T lymphocytes passed filtering and quality control, with 1,552 flow-sorted Tregs. After clustering and Treg identification (**Figure 14A**,**B**), we performed the same Wilcoxon rank sum test for differential gene expression comparing tumor-infiltrating Tregs *versus* peripheral blood Tregs (**Figure 14C**). Defining differential genes as adjusted P-value < 0.05 and log2-fold change < -1 or > 1, we found a total of 273 genes in our ccRCC-infiltrating Tregs compared to 467 genes in the HCC-infiltrating Tregs (**Figure 13F**). A total of 143 differential genes overlapped between Tregs infiltrating

into HCC and ccRCC tumors. Of note, we corroborated several previous reports on upregulation of *CCR8, LAYN,* and *MAGEH1* in tumor-infiltrating Tregs in both data sets (**Figure 13F,G**).[307–309] In addition, we observed increase expression of other Treg markers, including *CTLA4*, *ICOS*, *TNFRSF18*, *TNFRSF4*, and *TNFRSF9* (**Figure 13F,G**). The gene with the highest average log2-fold change in ccRCC (log2-fold change = 4.86) and HCC (log2-fold change = 4.55) tumor-infiltrating Tregs was *CD177*. A similar upregulation Treg-related genes, including *CD177*, was also seen in RNA-sequencing data sets derived from pooled tumor-infiltrating Tregs *versus* peripheral-blood controls in breast, colorectal, and non-small cell lung cancers (**Figure 14D,E**).

*Treg heterogeneity indicates the transcriptional bifurcation of tumor-infiltrating Tregs*

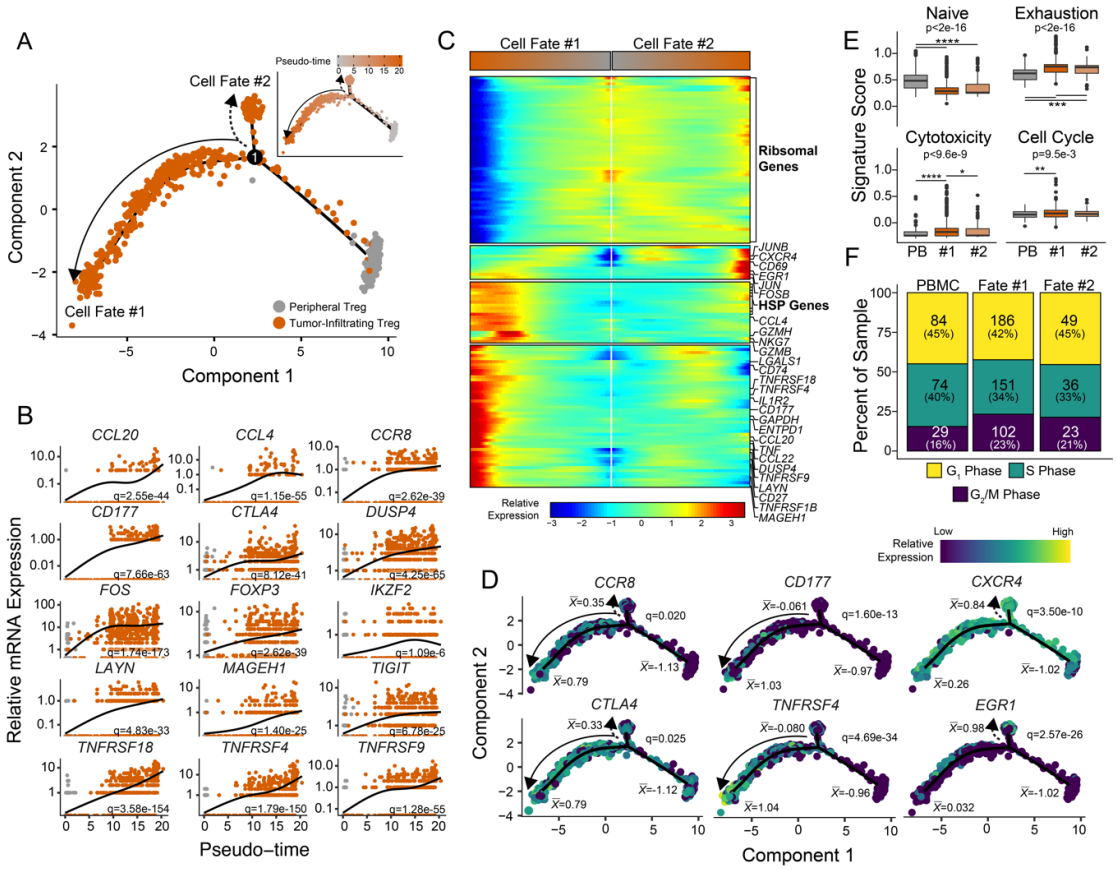


**Figure 15.** Transcriptional Heterogeneity in single-cell sequencing of ccRCC Tregs**. A.** Trajectory manifold of Tregs from the ccRCC using the Monocle 2 algorithm, solid and dotted line represent distinct cell trajectories/fates defined by single-cell expression profiles. **B.** Pseudo-time projections of transcriptional changes in immune genes based on the manifold. Significance based on differential testing based on the site of origin of Tregs was used to generate pseudo-time and adjusted for multiple comparisons. **C.** Expression heatmap of significant (Q < 1e-6) genes based on branch expression analysis comparing the two tumor-infiltrating cell fates and were used in the ordering of the pseudo-time variable. **D.** Cell trajectory projections of transcriptional changes in immune genes based on the manifold. Significance based on differential testing between the first and second cell fates of tumor-infiltrating Tregs. $\bar{x}$ denotes the scaled mean of each pole of the manifold. **E.** Gene signature analysis of the poles of the trajectory manifold. P-value based on one-way ANOVA with individual comparisons corrected for multiple hypothesis testing using the Tukey HSD method. * P < 0.05, ** P <0.01, *** P <0.001, **** P < 0.0001. **F.** Results of the cell cycle regression analysis of single cells for each cell fate using the Seurat R package.

Unlike these previous genomic studies that rely on pooled tumor-infiltrating in comparison to peripheral-blood Tregs, we also were able to investigate the dynamic transcriptomic processes of the Tregs at a single-cell level (**Figure 15A**). Using the machine-learning reverse graph embedding for dimensional reduction available in the Monocle 2 algorithm, we constructed a manifold using of the tumor-infiltrating and peripheral blood Tregs in the ccRCC data set (**Figure 15A**). This technique orders the single cells by expression profiles to represent distinct cellular fates or biological processes. Unlike differential gene expression, the ordinal construction of the manifold creates a pseudo-time variable that allows us to investigate the relative expression of genes in Tregs during the infiltration process (**Figure 15A,B**). Using the pseudo-time created by the reverse graph ordering, we also found a number of genes with significant role in the manifold ordering, including many previously identified in the differential expression analysis that are increasing in tumor-infiltrating by predicted pseudo-time projection (**Figure 15B**). In addition, we observed a significant increase in chemokine, *CCL20* and *CCL4*, over purported pseudo-time (**Figure 15B**). These chemokines are reported to play a role in the trafficking of Tregs to sites of antigen presentation.[322]

Next, to understand the branching structure of the manifold, we performed branched expression analysis modeling using the pseudo-time variable in the reduced model (**Figure 15C**). Commonly-associated markers of immune regulation and suppression had increased expression in tumor-infiltrating Tregs of Cell Fate #1 compared to the Tregs of Cell Fate #2 (**Figure 15C**). In contrast, the Tregs of Cell Fate #2 had maintenance of ribosomal-associated gene expression (**Figure 15C**). Using the differential gene analysis between the branches, we were able to see three distinct trends in gene expression between the tumor-infiltrating Tregs: non-specific increase between the two tumor-infiltrating Treg fates, like *CCR8* and *CTLA4*; increased in the Cell Fate #1, like *CD177* and *TNFRSF4*; and increased in the Cell Fate #2, like *CXCR4* and *EGR1* (**Figure 15D**). To investigate the potential functional difference between the cell fates, we performed gene signature analysis comparing single-cell gene expression of the poles of the manifold (**Figure 15E**). In both tumor-infiltrating Tregs cell fates, we observed a significant reduction in the naïve T cell signatures and an increase in T cell exhaustion signatures compared to peripheral-blood ccRCC Tregs (**Figure 15E**, upper row). The gene signature for cytotoxicity was significantly higher in Tregs of the first cell fate compared to peripheral blood or the second cell fate Tregs (**Figure 15D**). Similarly, we found an increase in signature scores for cell cycle in Tregs of the Cell Fate #1 (**Figure 15E**). Performing a regression-based analysis of cell-cycle-related genes, we assigned cell phases to each Treg, finding an increased percentage of $G_2M$ phase Tregs in in the first cell fate (**Figure 15F**). In addition to the single-cell RNA sequencing, we also sequenced the VDJ regions of the T-cell receptor (TCR). Examining the shared clonotypes, the matching TRA and TRB chains of the TCR across Tregs, we found nearly 0 shared sequences in the peripheral blood of each ccRCC (**Figure 16A**, upper panel), with modest increases in repeated clonotypes in tumor-infiltrating Tregs (**Figure 16A**, lower panel). Moving from peripheral-blood to tumor-infiltrating Tregs there is a significant increase in the percentage of clones, both

shared between two Tregs (**Figure 15B**) and three or more Tregs (**Figure 16C**). Combining the peripheral-blood and tumor-infiltrating Tregs by patient identity to the transcriptional manifold, we observed a significant increase in expanded clonotypes in the first tumor-infiltrating Treg cell fate (**Figure 16D**).

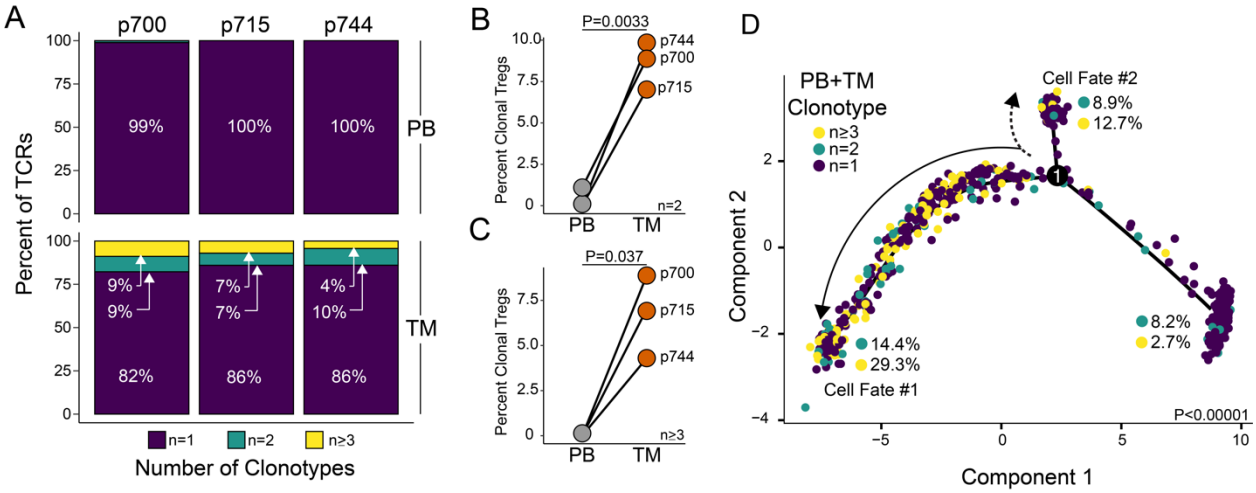


**Figure 16.** Clonotype analysis of ccRCC-infiltrating and peripheral-blood Tregs**. A.** Percentages of assigned clonotypes by patients in peripheral-blood (PB, upper panel) and tumor-infiltrating Tregs (TM, lower panel). **B.** Relative increase in clonotypes with two copies in the same patient comparing peripheral-blood (grey) to tumor-infiltrating (orange). Significance testing utilized T test with Welch's correction **C.** Relative increase in clonotypes with three or more copies in the same patient comparing peripheral-blood (grey) to tumor-infiltrating (orange) Tregs. Significance testing utilized T test with Welch's correction. **D.** Cell trajectory projections of shared clonotype percentages by cell fate, combining the peripheral-blood and tumor-infiltrating clonotypes of a single patient. Significance based on $\chi^2$ testing.

In order to corroborate our findings, we utilized the HCC single-cell dataset, finding a similar bifurcated architecture in the cell trajectory manifold of Tregs (**Figure 17A**). In addition, we noted a similar increase in immune modulatory genes were seen across pseudo-time in HCC-infiltrating Tregs (**Figure 17B**), however the contrast in expression values between tumor-infiltrating *versus* peripheral-blood Tregs was less for *CCR8*, *CTLA4*, *DUSP4*, *FOXP3*, *IKZF2*, *MAGEH1*, *TIGIT*, and *TNFRSF18* (**Figure 17B**). Within HCC-infiltrating Tregs, *CCR8* and *CTLA4* failed to differentiate between the two manifold-based Treg cell fates while *CD177* and *TNFRFS4* was still associated with Tregs of the first cell fate (**Figure 14C**). Additional markers specific for Tregs in Cell Fate #1 included: *HSPB1*, *CCL20*, *IFNG*, and *HLA-DRA* (**Figure 17C**) also seen in the renal-infiltrating Tregs of the first cell fate (**Figure 15C**). Similar to our ccRCC Tregs, we found a reduction in naïve T cell signatures and an increase in exhaustion gene signatures in the tumor-infiltrating populations (**Figure 17D**). Interestingly, the Tregs of the second cell fate appeared to have higher cytotoxicity and cell cycle signatures compared to peripheral blood Tregs and Tregs from the first cell fate (**Figure 17D**,**E**). This may be a result of creating the cell trajectory manifold from Tregs from different regions of the liver, in which may vary in expression and proliferation. In our combined examination of heterogeneity in tumor-infiltrating Tregs, we observed increased expression of *CD177* in ccRCC-infiltrating Tregs compared to peripheral blood Tregs

(**Figure 13G**) and the specificity towards a suppressive cell fate (**Figure 14D** and **Figure 17C**) leading us to further investigate the functional impact of CD177+ Tregs.

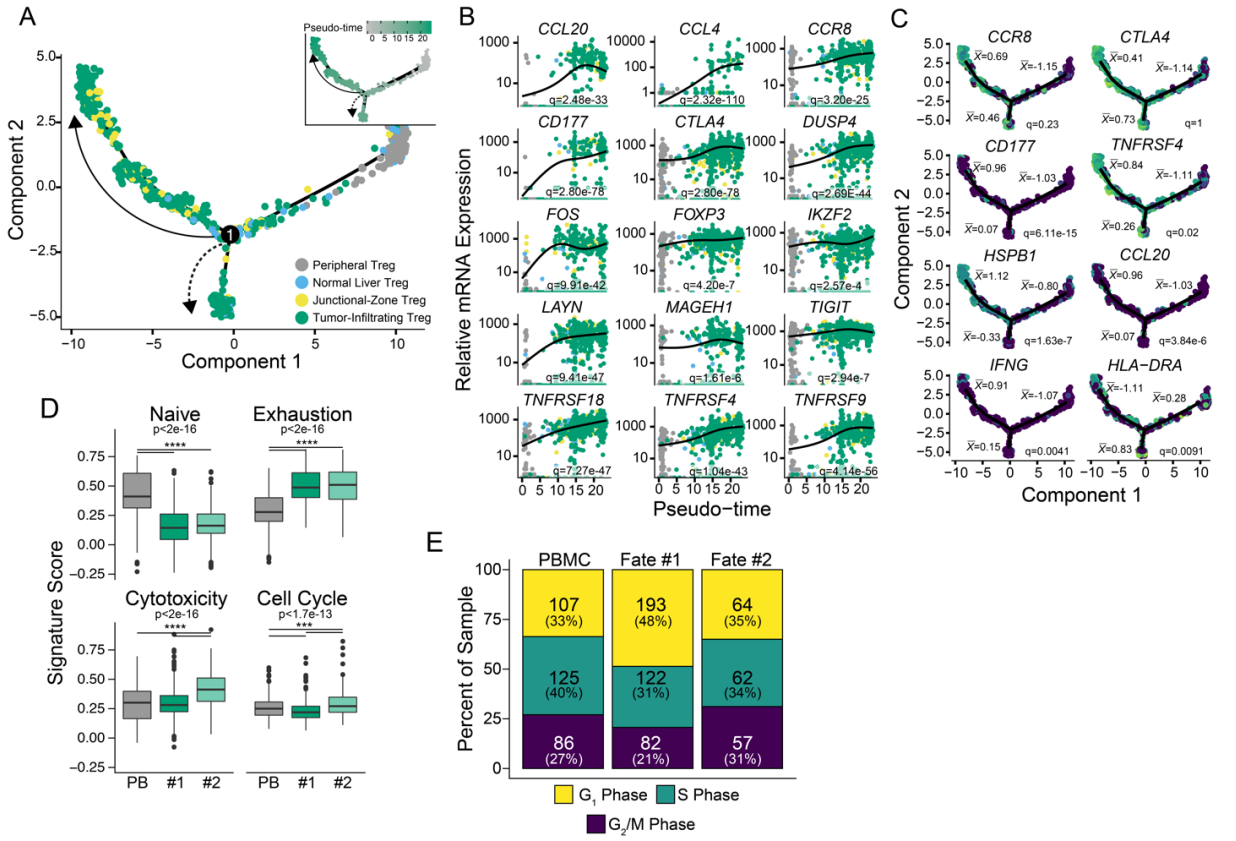


**Figure 17.** Transcriptional heterogeneity in single-cell sequencing of HCC Tregs. **A.** Trajectory manifold of Tregs from the HCC using the Monocle 2 algorithm. **B.** Pseudo-time projections of transcriptional changes in immune genes based on the manifold. **C.** Cell trajectory projections of transcriptional changes in immune genes based on the manifold. $\bar{x}$ denotes the scaled mean of each pole of the manifold. **D.** Gene signature analysis of the poles of the trajectory manifold. *** P <0.001, **** P < 0.0001. **E.** Results of the cell cycle regression analysis of single cells for each cell fate using the Seurat R package.

*CD177 defines a subset of suppressive tumor-infiltrating Tregs*

Our interest in CD177 is long standing and we have previously shown that CD177 plays a role in neutrophil viability using a genetic mouse model.[312] To corroborate our single-cell RNA findings of tumor-infiltrating Treg expression of CD77, we developed immunohistochemical (IHC) staining procedure for CD177 (**Figure 18A**). In general, we observed a paucity of staining for CD177 in tumor cells, but high levels of CD177 in tumor-infiltrating lymphocytes (TIL, **Figure 18A**). In collaboration with the Comparative Pathology Core in the Department of Pathology we performed dual IHC staining for FoxP3 (brown) and CD177 (red) and demonstrated a subset of CD177+ FoxP3+ T cells (**Figure 18B**, orange arrow). In addition to IHC, we performed flow cytometry on TILs in a number of cancer types find the CD177+ infiltrating population to be seen exclusively in CD4+ T cells (Figure 3C). Staining on CD177+ TILs demonstrated that majority of CD177+ cells were Tregs as identified by flow cytometry (**Figure 18D**, $CD3^+CD4^+CD25^+CD127^{low}$) as well as by intracellular FoxP3 staining (**Figure 18E**, $CD3^+CD4^+FoxP3^+$).

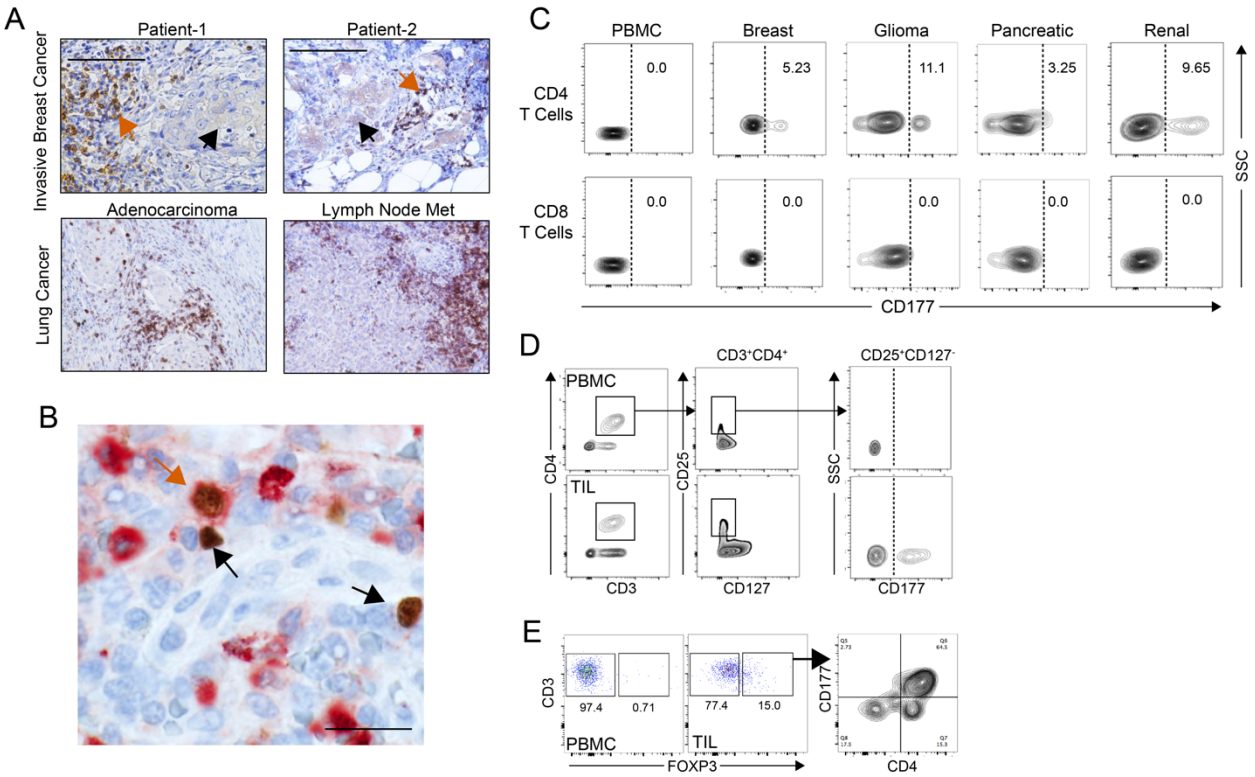


**Figure 18.** CD177+ tumor-infiltrating lymphocytes are comprised of Tregs. **A.** Human breast cancer, lung adenocarcinoma, and lymph node metastasis IHC for CD177 (brown) expression. **B**. Representative dual IHC staining for CD177 (red) and FoxP3 (brown) in breast carcinoma section. Dual-positive cells are indicated with orange arrows and FoxP3-positive cells are indicated by black arrows. **C.** Representative surface expression of CD177, as measured by flow cytometry, by CD4+ and CD8+ TILs isolated from different human cancer samples as well as PBMCs from a healthy individual. **D**,**E.** Representative flow cytometry data gating on Tregs identified as CD25+CD127- (**D**) or FoxP3+ (**E**) and isolated from TILs in breast cancer tissue or PBMCs.

Within the suppressive Cell Fate #1 Treg subset, we observed a significant increase in the number of CD177+ Tregs (**Figure 19A**,**B**). In order to support our single-cell RNA findings of TI Treg specificity for CD177, we performed flow cytometry on TILs. CD177+ Tregs consisted of 0 to 56.3% (median = 22.4%) of total Tregs in breast cancer and 6.8 to 48.1% (median = 16.8%) in renal cancer (**Figure 19C**), mirroring the roughly 15-25% of tumor-infiltrating Tregs in both the ccRCC and HCC datasets. There were small percentages of conventional CD4+ T cells (Tconv) expressing CD177 with a median of 0.31% in breast and 0.72% in renal tumors (**Figure 19C**). We observed that CD177+ Tregs had larger pool of cells expressing PD-1, CTLA-4 and CCR8 (27.5%, 33.3%, and 75% respectively) when compared to CD177- Tregs (6.2%, 0%, and 40% respectively), suggesting an active and suppressive phenotype (**Figure 19D**). CD27, T cell activation marker, exhibited similar pattern for CD177+ Treg or Tconv cells (52.7% vs 59.5%), but was completely lost in CD177- Tregs (**Figure 19D**). We were also able to isolate a small fraction of CD4+ conventional T cells (Tconv) with CD177 and the analysis of the markers, with the exception of CD27, did not show different patterns compared to CD177- Tconv cells (**Figure 19D**), indicative of a Treg-specific phenotype related to CD177 expression.

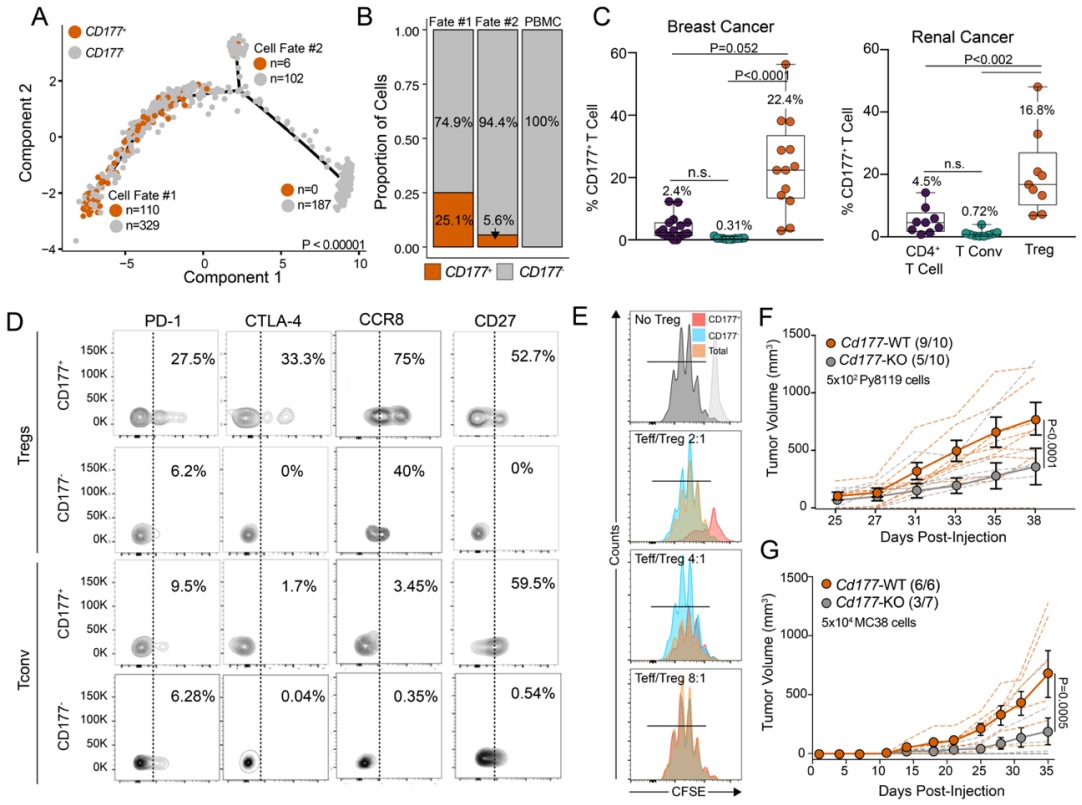


**Figure 19.** CD177+ Tregs are a suppressive subset of Tregs. **A**. Trajectory manifold of Tregs from the ccRCC Tregs with the number of *CD177*+ and *CD177*- Tregs for each respective cell fate. Significance based on $\chi^2$ testing comparing the three poles of the manifold. **B**. Proportional distribution of *CD177*+ Tregs by cell fate across the manifold. **C**. Percentages of CD177+ cells relative to different CD4+ TILs in breast cancer (n=13-18) and in renal cancer (n=8). Data are presented as mean ± SD with median values presented in the figure. **D**. Representative flow cytometry data for select marker expression by CD177+ or CD177- Tregs ($CD3^+CD4^+CD25^+CD127^{low}/FoxP3^+$) or Tconv ($D3^+CD4^+CD25^-/FoxP3^-$) isolated from breast cancers. n=2-3. **E**. Tumor-infiltrating Treg and T effector cells were isolated from breast tumors. Tregs were further sorted by CD177 expression and incubated with effector cells at indicated ratios. **F**. Py8199 tumor growth is significantly reduced in *Cd177*-KO mice compared to WT, P< 0.0001 (two-way ANOVA) in mice challenged with 500 cells per inoculation, n=10 bilateral tumors. Numbers in parenthesis equates to the number of nonrejected tumor challenge/total. **G**. MC38 tumor growth is significantly reduced in *Cd177*-KO mice compared to WT, P=0.0005 (two-way ANOVA) in mice challenged with 50,000 cells per inoculation. Numbers in parenthesis equates to the number of nonrejected tumor challenge/total. Data are presented as mean ± SEM. *In vivo* tumor modeling was performed by Dr. Ryan Kolb.

In order to confirm the immunosuppressive property of CD177+ Tregs, we performed an *in vitro* immunosuppression assay (**Figure 19E**). Interestingly, when non-sorted Tregs (include CD177+/- Tregs in their naturally occurring ratios) were employed, total Tregs were more suppressive than purified CD177- Tregs despite the fact that the immunosuppressant effect of CD177+ Treg was diluted by the presence of CD177- Tregs (**Figure 19E**). The scarcity of CD177+ tumor infiltrating Tregs made it challenging to repeat these experiments. However, given that the experiments were carried out using Tregs pooled from 3 individual breast cancer tissues, results can be considered biologically representative. We used *Cd177* KO mice and their WT littermate controls to verify the suppressive phenotype of CD177+ Treg. After titrating injections to identify optimal cell numbers to facilitate immune-based tumor control of Py8119 cells, a triple-negative breast cancer model, we settled on $5x10^3$ cells for injection. We found a clear phenotype in regard to tumor growth in KO *versus* WT mice, where KO mice rejected tumor inoculation (5 out of 10 total compared to 1 out of 10 WT)

and showed significantly slower tumor growth (P<0.0001) (**Figure 19F**) pertaining to immune-mediated tumor clearance at such low inoculation dose. The same increase in tumor rejection (4 of 7) and slower tumor growth (P=0.0005) was seen in the syngeneic grafts of MC38 colorectal cells in *Cd177*-KO *versus* WT controls (**Figure 19G**).

*Superior prognostic ability of suppressive Treg subset gene signature*

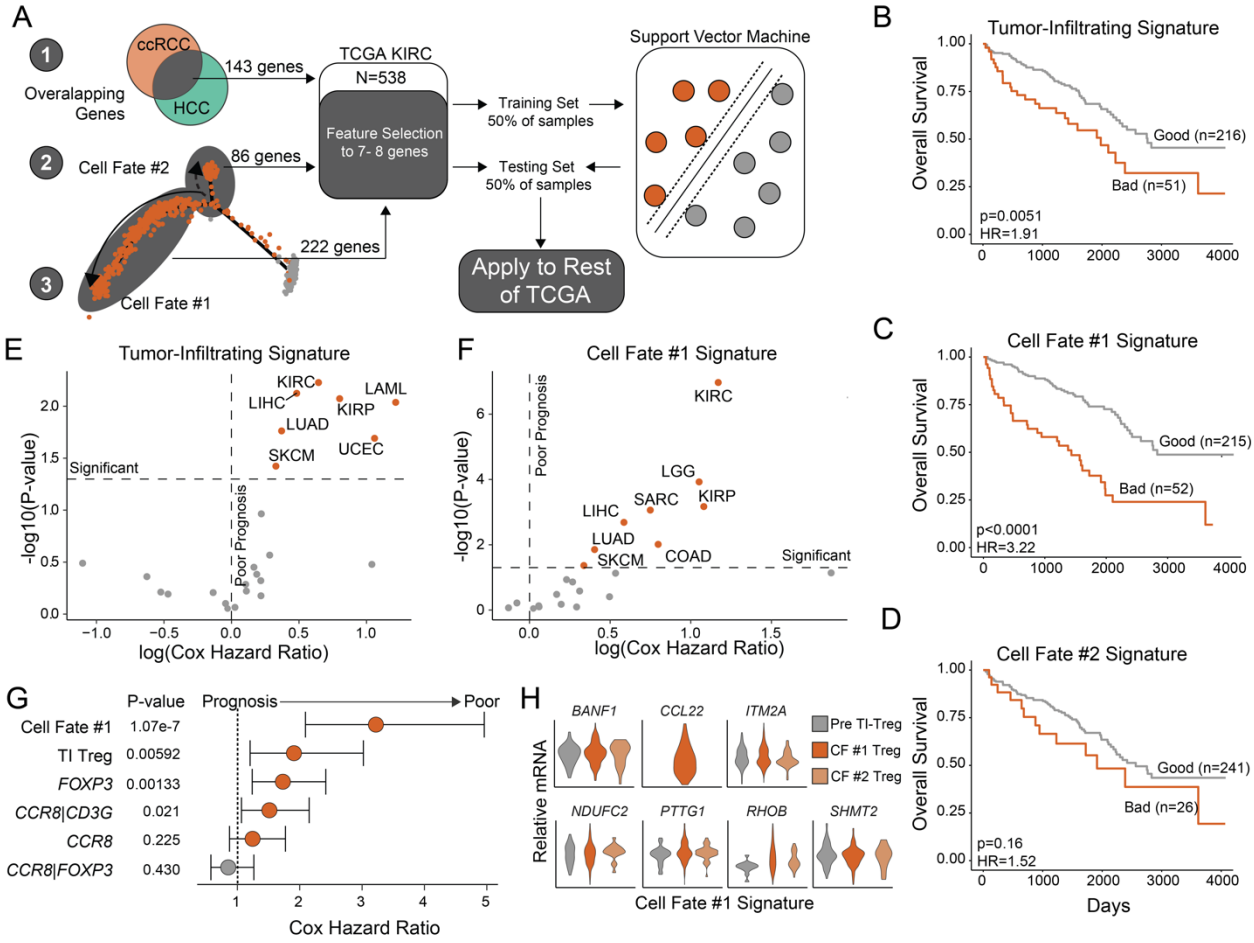


**Figure 20.** Improved prognostic prediction associated with suppressive Tregs. **A**. Schematic of signature development using feature selection from: 1) 143 differential genes of tumor-infiltrating Treg in ccRCC and HCC, 2) 86 genes differentially expressed in Cell Fate #2, and 3) 222 genes differentially expressed in Cell Fate #1 using the Kidney Renal Clear Cell (ccRCC, n=538) dataset from the TCGA. **B**. Kaplan-Meier curve for overall survival in ccRCC using the TI gene signature. **C**. Kaplan-Meier curve for overall survival in ccRCC using the Cell Fate #1 Treg gene signature. P-value based on log-rank test and hazard ratio (HR) based on Cox proportional hazard regression. **D**. Kaplan-Meier curve for overall survival in ccRCC using Cell Fate #2 Treg gene signature. **E**. Overall survival prediction with Cox proportional hazard ratio and -log10(P-value) based on log-rank testing across the 24 largest TCGA datasets using the tumor-infiltrating Treg signature. **F**. Overall survival prediction with Cox proportional hazard ratio and -log10(P-value) based on log-rank testing across the 24 largest TCGA datasets using the Cell Fate #1 signature. **G**. Prognostic prediction for single-cell Treg signatures compared to other proposed signatures for TI Tregs. Hazard ratios, 95% confidence intervals, and P-values derived from Cox proportional hazard regression modeling. **H**. Relative mRNA violin plots of the Cell Fate #1 signature based on the transcriptional trajectory state across the 160 PB Tregs and 574 TI Tregs.

As our data suggests the transcriptional and functional difference in a suppressive subset of tumor-infiltrating Tregs across several cancers, we predicted the gene signature development from single-cell data would

provide improved prognostic ability. We performed feature selection to identify gene signatures on: 1) 143 differential genes of tumor-infiltrating Treg in ccRCC and HCC, 2) 86 genes differentially expressed in the Cell Fate #2, and 3) 222 genes differentially expressed in the Cell Fate #1 using the Kidney Renal Clear Cell (ccRCC, n=538) dataset from the TCGA (**Figure 20A**). Using 50% (n=269) of the KIRC samples as a training set, we utilized supervised machine learning using a linear support vector machine model. After evaluating polynomial support vector machines, k-nearest neighbors, boosted tree classifications, and bootstrap aggregating classifications supervised machine learning models, we selected a linear support vector machine because the method had the least issues in the overfitting of survival data. Applying these signatures to the remaining 50% of KIRC dataset, we show the tumor-infiltrating signature (**Figure 20B**) and the Cell Fate #1 signature (**Figure 20C**) could discriminate between patients with good and poor prognosis, but not the Cell Fate #2 signature (**Figure 20D**). Despite both the tumor-infiltrating signature and Cell Fate #1 signature significantly predicting poor overall survival in roughly 20% patients, the Cell Fate #1 signature had superior ability to discriminate prognostic groups with a Cox hazard ratio of 3.22 compared to 1.91(**Figure 20B**,**C**).

The superior discrimination of the Cell Fate #1 Treg signature compared to the tumor-infiltrating signature was not isolated to ccRCC. We applied the signatures across the 24 largest TCGA datasets, finding the tumor-infiltrating signature significantly separating prognostic groups in 7 cancer datasets with P-values ranging from 0.006 to 0.038 and Cox hazard ratios ranging from 1.39 to 3.37 (**Figure 20E**). In contrast, the Cell Fate #1 Treg signature discriminating prognostic groups in 8 cancer types, with P-values ranging from 1.07e-7 to 0.043 and higher Cox hazard ratios of 1.4 to 5.5 (**Figure 20F**). The Cell Fate #2 Treg signature failed to discriminate groups based on overall survival across all of the 24 TCGA datasets (data not shown). Both the Cell Fate #1 Treg signature and the tumor-infiltrating Treg signature separated prognostic groups in checkpoint-inhibitor responsive melanoma (SKCM) and lung adenocarcinoma (LUAD), as well as renal papillary cell carcinoma (KIRP) and liver hepatocellular carcinoma (LIHC) (**Figure 20E**,**F**). In addition, Cell Fate #1 Treg signature separated prognostic groups in sarcoma (SARC) , low-grade glioma (LGG) and colon adenocarcinoma (COAD), the latter of which has had reports of Treg infiltration as being good prognostic indicators.[274] Within the KIRC dataset, we compared previously identified Treg signatures to our single-cell signatures, demonstrating superior prognostic ability in univariate Cox proportional hazard regression analysis (**Figure 20G**), In addition to investigating the prognostic ability of the Cell Fate #1 Treg signature, we also show that the signature is not exclusive to the first cell fate, but rather has higher mean expression and range of the seven genes selected: *SHMT2*, *CCL22*, *ITM2A*, *BANF1*, *PTTG1*, *RHOB*, and *NDUFC2* (**Figure 20H**).

**Discussion**

Due to the immunosuppressive role of Tregs in peripheral tolerance, the targeting for Tregs for therapy is a double-edged sword. While the attenuation of the Treg-mediated suppressive activity increases the anti-tumor immune response, Treg dysfunction is associated with autoimmune and inflammatory diseases. For example, CTLA-4 is constitutively expressed on Tregs and acts as a primary mechanism of immune suppression by Tregs, acting to block co-stimulation of T effector cells.[68] The anti-CTLA-4 antibody, ipilimumab, inhibits Treg function by abolishing CTLA-4-mediated suppressive signaling and depletes Tregs in the tumor microenvironment.[297–299] However, the whole body abrogation of Treg function leads to substantial side effect profile, including autoimmune and inflammatory conditions, like colitis, which can occur in over 50% of patients.[259,296] Thus, refining biomarkers for tumor-infiltrating Tregs is vital for the development of better targeted immunotherapy.

From our large single-cell RNA-sequencing cohort, we isolated 160 peripheral-blood Tregs and 574 tumor-infiltrating Tregs, representing one of the largest collections of expression profiles available for paired Treg samples. We expanded our analysis to include flow-sorted Tregs derived from HCC tumors with 264 peripheral-blood and 634 tumor-infiltrating Tregs.[309] We found a common differential signature of 143 genes comparing tumor-infiltrating *versus* peripheral-blood Tregs (**Figure 13F**). The overlapping genes included common Treg markers, like *CTLA4*, *ICOS*, *TNFRSF4*, *TNFRSF9*, *TNFRSF18*, as well as previously identified tumor-infiltrating Treg markers, like *CCR8*, *LAYN*, and *MAGEH1* (**Figure 13G**).[296,307–309] However, among the most consistently elevated genes in both single-cell RNA sequencing (**Figure 13G**) and pooled RNA sequencing (**Figure 14E**) of tumor-infiltrating Tregs was *CD177*.

CD177 has almost exclusively been studied in neutrophils, where it has been suggested to play a role in transendothelial migration, cell viability, and bactericidal activities.[312,323] Apart from a minor point in a recent paper on tumor-infiltrating Tregs in breast cancer in which the authors found that CD177 was expressed on a small subset of tumor-infiltrating Tregs[308] and a paper showing that its expression is a good prognosis marker in CRC,[314] little is known of CD177 expression and function outside of neutrophils. Across single-cell sequencing and flow cytometry, we observed CD177 in 15-25% of tumor-infiltrating Tregs in various (**Figure 13C**, **Figure 14C**, and **Figure 19A**,**C**). Analysis of transcriptional heterogeneity of tumor-infiltrating Tregs revealed CD177 closely associated with a more suppressive cell fate with elevated immune suppression markers (**Figure 15C**). Flow cytometry also indicated that the CD177$^+$ Treg population has higher number of cells with checkpoint proteins, CTLA-4 and PD-1, as well as suppressive markers CCR8 and CD27, compared to the CD177$^-$ population further supporting that CD177$^+$ Tregs are more suppressive than CD177$^-$ Tregs (**Figure 19D**). This increased level of suppression mediated by CD177$^+$ Tregs was overserved in *in vitro* suppression assays (**Figure 19E**) and *in vivo* tumor models (**Figure 19F**,**G**). Taken together, our work suggests CD177 as a

specific marker for a suppressive subset of Tregs. However, the specific function of CD177 in Tregs, how CD177 expression in Tregs is regulated, and whether CD177 can be used as a target for immunotherapy warrants further investigation.

Identifying specific markers and function of different tumor-infiltrating Treg subsets is important in both understanding the role of Tregs in immunotherapy, predicting prognosis or response of a patient to immunotherapy, and the development on new therapies. In the context of prognosis, *FOXP3* alone has had mixed results in the ability predict overall survival for patients.[274,302,303] More recently, two analyses of RNA-sequencing of tumor-infiltrating Tregs reported the use of the ratios, *CCR8/FOXP3* for breast cancer and *CCR8/CD3G* for lung and colorectal cancers, had modest predictive value in terms of overall survival.[307,308] The adequacy of single or double gene approach to survival prediction does not account for the observed heterogeneity within tumor-infiltrating Tregs (**Figure 16**). Similar to algorithms developed to quantify the immune contribution of bulk RNA sequencing,[267] we utilized linear support vector machines to develop a signature. The cell fate #1 Treg signature that predicted overall survival in 8 cancer types, in immunogenic cancers, like KIRC/ccRCC, SKCM and LUAD (**Figure 20F**). Surprisingly, the Cell Fate #1 signature discriminated prognostic groups in COAD, a cancer type with previous reports of Treg infiltration being a marker of good prognosis.[274] Although the Cell Fate #1 signature did not include *CD177*, likely a result of epithelial expression of *CD177* (**Figure 18A**) and the tumor-intrinsic role of CD177 as a tumor suppressor (Kolb, R., and Kluz, P. *et al* 2018, publication forthcoming), it did discriminate between transcriptional states based on our manifold (**Figure 20G**,**H**).

In summary, the Treg tumor infiltrate quantity and quality, including the presence of the suppressive Treg subset, are key to improved therapeutic and prognostic outcomes. Our study highlights the potential of single-cell genomics to improve the understanding of the tumor microenvironment and identify new targets of Treg-based immunotherapy. This data also provides a unique resource of transcriptome data from ccRCC peripheral-blood and tumor-infiltrating immune cells.

# CHAPTER 4
## SINGLE-CELL PROFILING OF CUTANEOUS T-CELL LYMPHOMA REVEALS UNDERLYING HETEROGENEITY ASSOCIATED WITH DISEASE PROGRESSION

### Rationale

Cutaneous T cell lymphomas (CTCLs) are a group of heterogenous T cell neoplasms with skin involvement. Two predominant types of CTCL include mycosis fungoides (MF) and Sézary syndrome (SS), both of which are thought to be derived from mature skin-homing CD4+ T cells.[324,325] Given this commonality and their often overlapping clinicopathologic features, MF and SS had historically been regarded as closely related entities on a spectrum; however, recent elucidation of distinct cells of origin[326] has favored MF and SS to represent distinct clinical entities.[327–329] SS refers to a rare form of CTCL characterized by circulating malignant cells with widespread skin involvement and possesses a poor 5-year survival rate.[324,330] In contrast, MF refers to a substantially more common CTCL with a skin-predominant, and usually a skin-limited presentation. MF most often has an indolent course, with a 5-year survival of 70-80%[328,330]; however, a subset of patients exhibit a progressive course such that malignant cells may be identified in the circulation, lymph nodes, and viscera. Treatments for advanced stage MF and SS ultimately become ineffective, contributing to the morbidity and mortality of this patient population. Methods to identify those patients who will progress to advanced and widespread disease may facilitate optimal transition from skin-directed therapies to more aggressive treatment, but such methods have not yet been established.

Despite a number of high-quality computational inquiries into the genomic makeup of CTCL[331–335], the development of differentiated T cells phenotypes and their relationship to disease pathogenesis represents a knowledge gap in the understanding of CTCL. In particular, the contribution of Treg-like cells to the malignant population in MF/SS has been controversial, with heterogeneous and sometimes conflicting results.[336–340] Heterogeneity within SS has been suggested by a recent targeted gene sequencing of single cells.[341] A deeper understanding of differences within the clonal malignant population in CTCL may yield insights into more effective treatment regimens and strategies.

Here, we use single-cell RNA sequencing and single-cell V-D-J sequencing to examine SS at a previously unrealized transcriptomic resolution by pairing isolated SS cells with matched normal CD4+ T cells. Using this unique dataset, we investigated the degree, as well as trajectory, of heterogenous transcriptional profiles within the malignant cell population to identify novel markers of SS that may aid in the detection, diagnosis, and staging of CTCL. We further validate the power of our methodology by applying our findings to a publicly available dataset consisting of a large cohort of CTCL patients and demonstrate that when used in conjunction

with an artificial intelligence (AI)-based approach, transcripts can be identified that distinguish early and late stage disease. A summary of the process and results can be found in **Figure 21**.

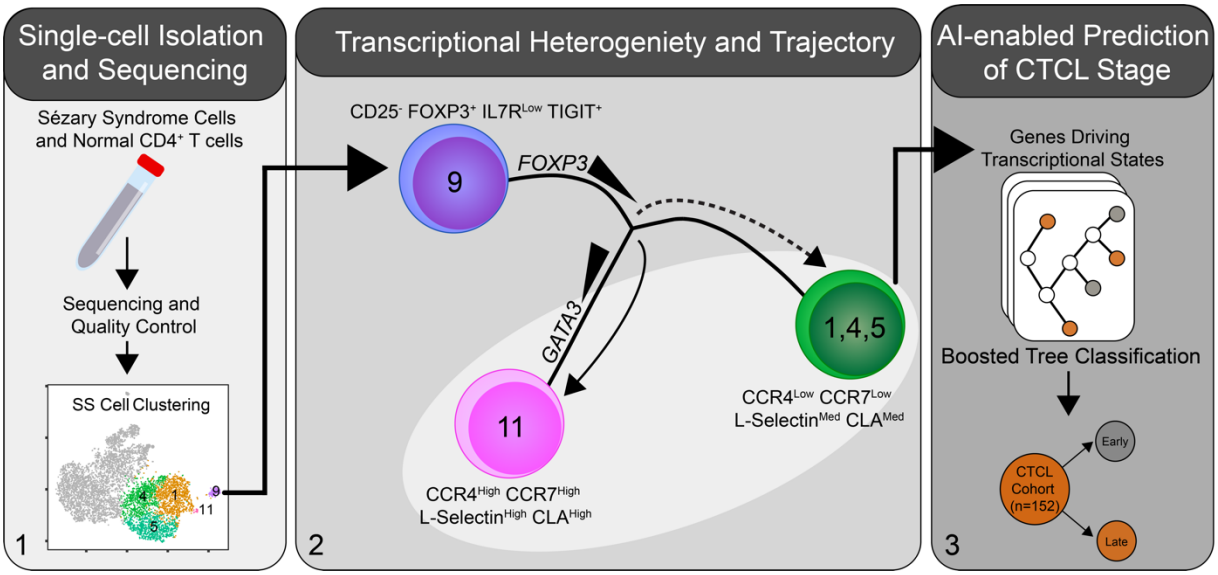


**Figure 21.** Basic scheme of analysis of single-cell SS sequencing. **1.** Single cell isolation, sequencing, and clustering of peripheral blood and SS cells. **2.** Transcriptional manifold construction of malignant SS revealed a progression from Treg-like cells into central memory T cells. **3.** Using the significant genes from pseudo-time, construction of a boosted tree classification predicted CTCL disease state with 79.6% accuracy.

## Materials and methods

*Patient recruitment*

The current study was approved by the University of Iowa Institutional Review Board and conducted under the Declaration of Helsinki Principles. The patient was recruited from the University of Iowa Cutaneous Lymphoma clinic in the Department of Dermatology. Informed written consent was received from the participant before inclusion in the study.

*Flow cytometry*

A blood draw was performed, and peripheral blood mononuclear cells (PBMCs) were isolated using a Ficoll gradient. Cells were labeled with fluorescent antibodies specific for CD3, CD4, CD8, CD45RA, CD45RO, CD5, CD7, and CD26 and flow sorted on a Becton Dickinson Aria II.

*Single-cell RNA sequencing*

A malignant ($CD3^+CD4^+CD5^{bright}SSC^{hi}$) and nonmalignant CD4 ($CD3^+CD4^+CD5^{int}SSC^{int}$) population were flow sorted in parallel. T-cell receptor sequencing and 5' gene expression sequencing was performed using the Chromium (10x Genomics, Pleasanton, CA) and Illumina (San Diego, CA) sequencing technologies. Amplified cDNA was used to construct both 5' expression and TCR enrichment libraries. Libraries were pooled together

and run on sperate lanes of a 150 based-paired, paired-end, flow cell using the Illumina HiSeq 4000. Basecalls were converted into FASTQs using the Illumina bcl2fastq software by the University of Iowa Genomics Division. FASTQ files were aligned to human genome (GRCh38) using the CellRanger v2.2 pipeline as described by manufacturer. Single-cell immune profiling of the clonotypes of the $CD4^+$ T cells was performed in conjunction with the single-cell RNA sequencing following the protocols described above.

*Single-cell data processing and analysis*

Initial processing of peripheral (n=4,485) and malignant (n=3,526) $CD4^+$ T cells was performed using the Seurat R Package (v2.3.4).[316] Individual cells were adjusted for total expression and percentage of mitochondrial reads. After processing, clustering was performed using the Seurat package on peripheral (n=4,436) and malignant (n=3,443) $CD4^+$ T cells. Dimensional reduction to form the tSNE plot utilized the top 10 calculated principle components and a resolution, or granularity of the clusters, of 1.2. Cluster markers and differential gene expression analyses were performed using the Wilcoxon rank sum test. Single-cell immune phenotyping utilized the SingleR (v0.2.0) R package on mean raw count data for clusters identified in Seurat.[342] Differential markers between peripheral and malignant $CD4^+$ T cells utilized the percentage of cells that express the individual mRNA species and the average log2-fold change between the two cell populations. Cell trajectory and pseudo-time analysis was performed using the Monocle R package (v2.8.0) and the reverse graph embedding machine learning algorithm.[318] Differential gene testing for the pseudo-time analysis was based on the previously identified malignant cell clusters and a cut-off for significance q-value < 0.01.

*Machine-learning gene signature analysis*

Raw TruSeq FASTQs of 152 CTCL and 29 normal/benign skin lesions were downloaded from SRP114956.[343] Additional clinical data on patient age and disease were downloaded from the SRA repository. Files were pseudo-aligned with kallisto using the GRCh38 build of the human transcriptome.[319] Transcript-level quantifications were condensed to gene-level and scaled to transcripts-per-million (TPM) values. In total, 344 genes were sequenced and quantified across the 151 of the total 152 patients, a single patient was removed due to alignment issues. Quantified genes were then cross-referenced to significant (q-value < 0.05) genes identified in the Monocle 2 algorithm to narrow down signature candidates, with a total of 93 genes used for classification prediction. Boosted classification trees were constructed with the gbm (v2.1.3) package using log TPM values. Boosting was performed with 10,000 classification trees with a multinomial distribution; interaction depth and shrinkage parameters were selected via 10-fold cross validation on the training dataset. Variable importance of each gene was assessed by quantifying the mean decrease in the Gini index of each predictor averaged over all splits.

## Results

*Differential Clustering of Malignant and Normal CD4+ T cells*

We performed parallel single-cell RNA sequencing and T-cell receptor (TCR) VDJ sequencing of sorted malignant CD4+ T cells paired with normal CD4+ T cells using a single cell droplet platform, as outlined in **Figure 22A**. Malignant CD4+ T cells were identified in a patient with SS by high side scatter[326] as well as aberrantly high expression of CD5; these cells made up nearly 86% of circulating CD4+ T cells. Normal CD4+ T cells were sorted in parallel, with a normal side scatter profile and normal CD5 expression (**Figure 22B**).

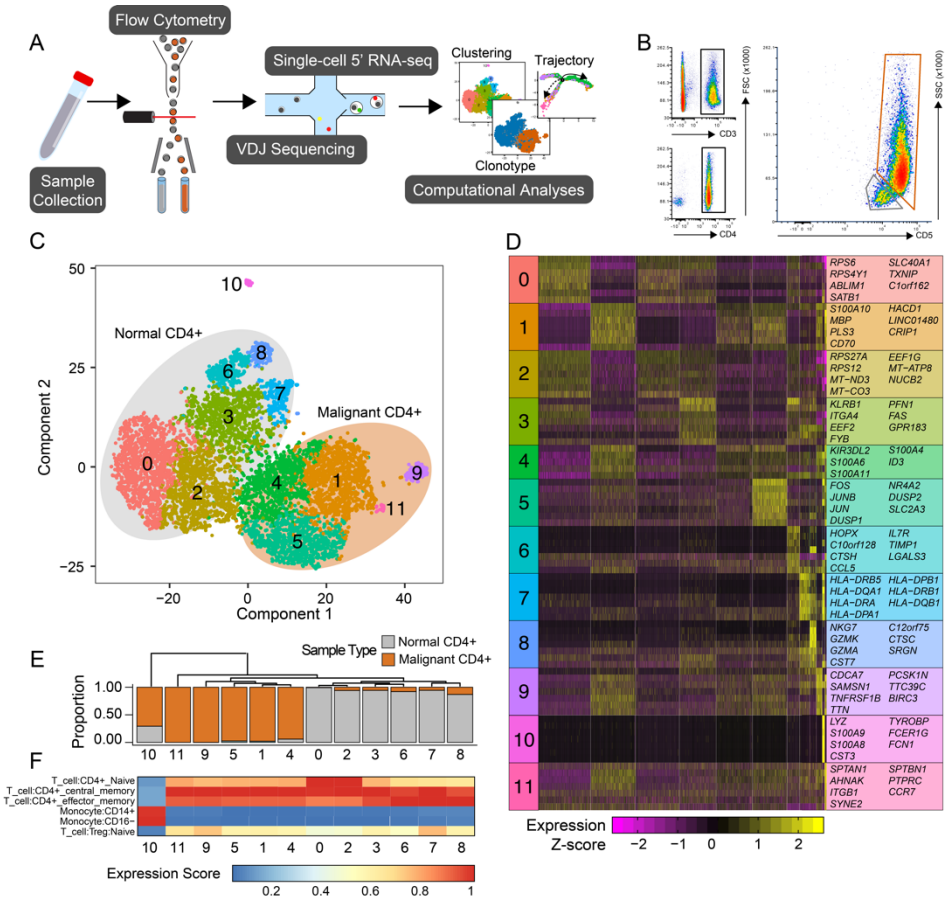


**Figure 22.** Single-cell isolation and sequencing of peripheral blood and Sézary Syndrome cells. **A.** Schematic of the isolation, sequencing, and analysis of the single-cells. **B.** Flow cytometry gating of the patient sample to isolate peripheral blood and tumor cells. **C.** tSNE projection of patient sample with normal peripheral blood samples (n=4,436) outline in grey and tumorigenic CD4 cells (n=3,443) in orange. **D.** Unique cluster markers based on the Wilcoxon rank sum test, adjusted P-value < 1e-50. **E.** Phylogenic tree of cluster identities based on mean mRNA values in the cluster with corresponding cluster composition. **F.** Normalized enrichment score of predicted immune cell phenotype based on SingleR algorithm for each tSNE cluster.

After single-cell RNA and TCR sequencing of the isolated CD4+ T cells, data were filtered for low-quality cells and normalized. Assessing the collective heterogeneity of both normal and malignant CD4+ T cells, we observed 12 distinct clusters based on mRNA expression (**Figure 22C**). Accompanying the clustering, we also identified the top 5-7 genes that define each cluster (**Figure 22D**). Of the tSNE clusters, 6 were comprised of normal CD4+ T cells, while 5 consisted of the malignant SS cells. Using Euclidean hierarchical clustering, we found the tSNE clusters were most closely related to the normal *versus* malignant classification (**Figure 22E**), further confirming the separation within the tSNE itself. Using the mean mRNA expression of each cluster, we correlated the gene expression with known marker genes, with the majority of cells of both normal and malignant origin correlating with CD4+ central memory T cells (Tcm, **Figure 22F**). Notably, the normal CD4+

T cell Clusters 0 and 2 appeared to contain a naïve CD4+ T cell phenotype, and the clusters corresponding to the malignant SS cell population appeared to contain a phenotype consistent with Tcm.[326] An additional cluster (Cluster 10) consisted of CD4+ myeloid cells and was excluded from further analysis.

*SS cells are clonal and transcriptionally distinct from normal CD4+ T cells*

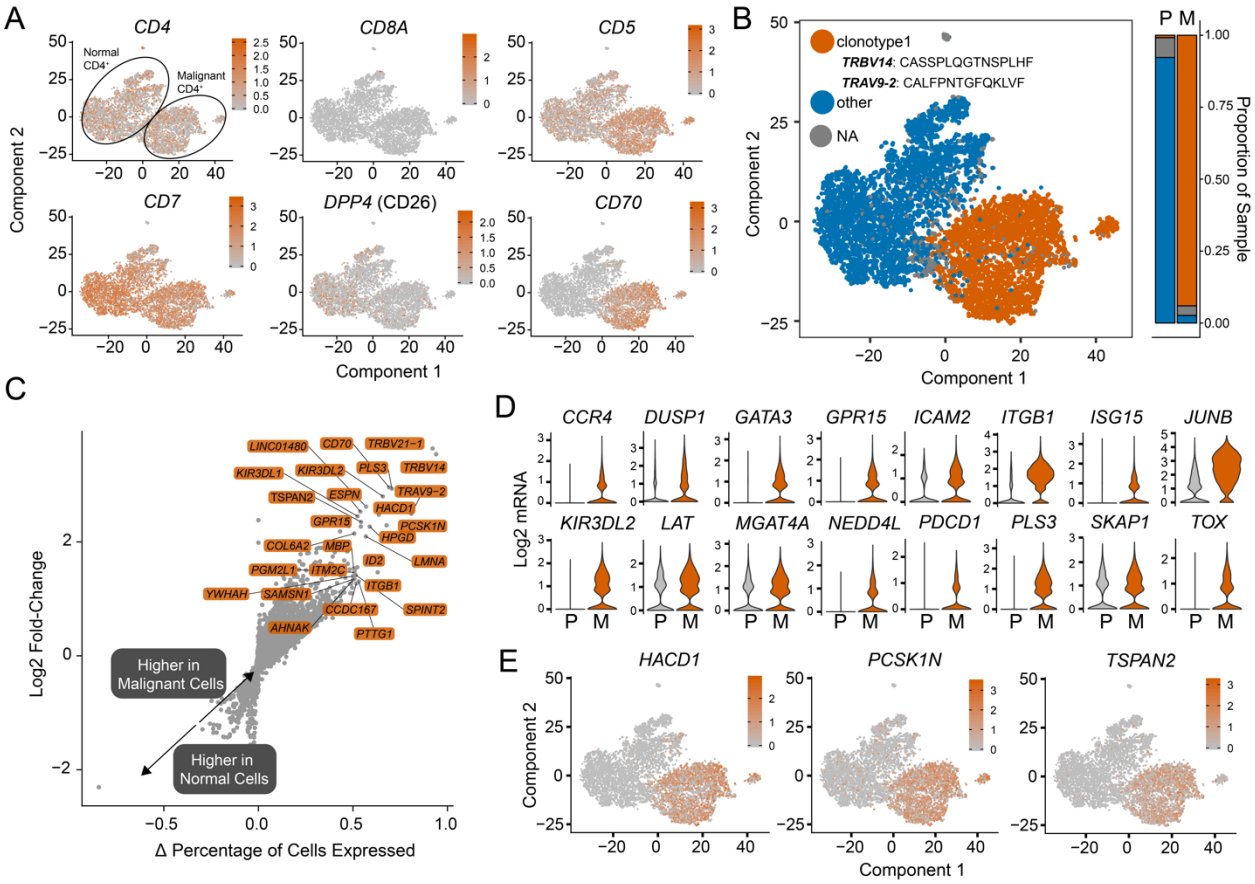


**Figure 23.** Transcriptomic comparison of malignant *versus* normal CD4+ T cells **A.** tSNE projects of common markers used to diagnose CTCL **B**. VDJ sequencing of malignant CD4+ T cells examining the distribution of a single prominent clonotype in the malignant T cells (orange) **C**. Log2-fold change expression *versus* the difference in the percent of cell expressing the gene comparing malignant to normal peripheral blood CD4+ T cells (Δ percentage of cells expressed). **D.** Violin plots of previously identified markers of CTCL (adjusted P < 1e-10). **E**. Potential novel markers of CTCL cells with a Δ percentage of cells expressed greater than 50% and adjusted p-values < 1e-100.

In order to investigate the difference in malignant and normal CD4+ T cells, we confirmed the separation of normal and malignant cells using previously identified markers (**Figure 23A**). We verified that sequencing was performed on isolated CD4+ T cells, and that the malignant population exhibited a characteristic decrease in CD26 (*DPP4*) and increase in *CD70* (**Figure 23A**). As previously mentioned, the patient's malignant cells expressed an aberrant increase in CD5, which we could demonstrate at the mRNA level (**Figure 23A**). To further demonstrate the observable difference in the malignant SS cells, we filtered the VDJ sequencing results for the top TCR hits and matched 94.7% of sequenced cells with the corresponding VDJ sequencing

information. Of the 3,328 cells sorted for the SS phenotype (**Figure 23B**) with recoverable TCR sequencing information, 97.3% consisted of a single clonotype containing *TRBV14* (CDR3 amino acid sequence: CASSPLQGTNSPLHF) and *TRAV9-2* (CDR3 amino acid sequence: CALFPNTGFQKLVF) (**Figure 23B**). In contrast, the normal CD4+ T cells had 4007 unique clonotypes with 37 individual cells (0.9%) possessing the same malignant *TRBV14/TRAV9-2* clonotype (**Figure 23B**), likely due to the close proximity of the flow sorting gates to each other.

In order to investigate potential novel markers or therapeutic targets of SS, we performed differential gene analysis comparing the malignant and normal CD4+ T cells. We used this comparison analysis by contrasting the log2-fold change (y-axis) and the difference in the percentage of cell expressing the gene (Δ percentage of cells expressed, x-axis) (**Figure 23C**). By examining the difference in the percentage of malignant *versus* normal peripheral blood CD4+ T cells, this allows for the identification of specific markers of SS. As expected, of the genes with the highest log2-fold change and greatest discrimination between malignant *versus* normal cells were *TRBV14* (log2-fold change = 3.53, Δ percentage = 94%) and *TRAV9-2* (log2-fold change = 2.49, Δ percentage = 81%) (**Figure 23B**). Interestingly the majority of malignant SS cells also expressed a second TRBV region variant, *TRBV21-4* (CDR3 amino acid sequence: CALFPNTGFQKLVF) (**Figure 23C**). Using this analysis, we examined previously identified genes that relied on pooled SS RNA sequencing and found differential expression in *CCR* [344], *DUSP1*[345,346], GPR15[347], *ICAM2*[346], *JUNB*[346,348], *KIR3DL2*[349], *PLS3*[350], *ITGB1*[333,345], *GATA3*[345,346], *NEDD4L*[332,347], *LA* [351], *MGAT4A*[351], *PDCD1*[333,351,352], *SKAP* [334,351], and *TOX*[351,353,354] (**Figure 23D**). We also identified novel markers, selecting three genes, *HACD1*, *PCSK1N*, and *TSPAN2* with log2-fold change > 2 and Δ percentage of cells expressed > 50% (**Figure 23E**).

*Heterogeneous transcriptional profiles of single cells in SS*

Unlike previous genomic studies that rely on pooled SS cells in comparison to normal CD4+ controls, we also were able to investigate the heterogeneity of the SS cells at a single-cell level (**Figure 24A**), and our previous analysis separated this malignant population into 5 clusters. Using the machine-learning reverse graph embedding for dimensional reduction available in the Monocle 2 algorithm, we constructed a manifold using the malignant SS cells (**Figure 24B**). This technique orders the single cells by expression patterns to represent distinct cellular fates or biological processes.[318] Despite our finding of the clonal expansion of the SS cells (**Figure 23B**), we observed distinct bifurcated architecture of the manifold (**Figure 24B**). Based on this ordering, SS cells appear to start principally from Cluster 9 and moved towards Cluster 11 (solid line) or Clusters 1,4 and 5 (dotted line) (**Figure 24B**).

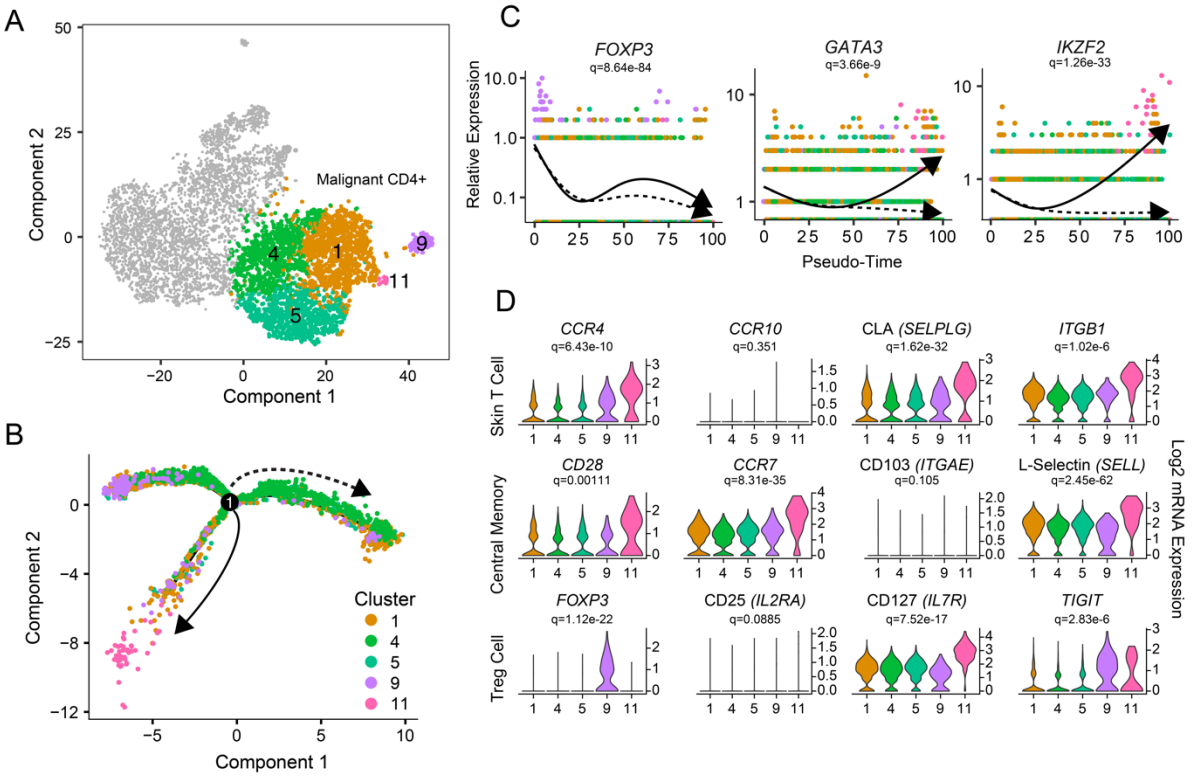


**Figure 24.** Transcriptional heterogeneity in malignant CD4+ T cells. **A.** tSNE projection of patient malignant CD4 cells (n=3,443). **B.** Trajectory of malignant cells from clusters 1, 4, 5, 9, and 11 using the Monocle 2 algorithm, solid and dotted line represent distinct cell trajectories defined by single-cell transcriptomes **C.** Pseudo-time projections of major immune transcriptional drivers in the malignant CD4+ T cells. Significance based on differential testing by cluster identification used to generate pseudo-time and adjusted for multiple comparisons. **D.** Selection of genes by cluster identity for skin-homing, central memory, and regulatory T cell phenotypes. Significance based on the pseudo-time generated by the Monocle 2 algorithm and correct for multiple comparisons.

In order to better understand the differential genes driving the ordinal construction of the manifold, we examined major immune transcription factors expression across the malignant CD4+ T cells, focusing on *FOXP3*, *GATA3*, and *IKZF2* (**Figure 24C**). Using the pseudo-time created by the reverse graph ordering, we observed a general decrease in *FOXP3* in both directions of the bifurcation (**Figure 24C**). In contrast, both *GATA3* and *IKZF2* (HELIOS) had marked increase expression with the cell fate associated with cluster 11 (**Figure 24C**). Utilizing the differential expression analysis based on the pseudo-time construction, we next investigated the underlying differences of the malignant clusters by defined immune phenotypes. We separated the analysis into markers of skin-homing T cells, central memory T cells, and Tregs (**Figure 24D**). As expected we found consistent expression of skin-homing markers, *CCR4*, *SELPLG* (CLA) and *ITGB1,* while there was a lack of expression of *CCR10* across all clusters (**Figure 24D**, upper row). Similarly, the malignant cells exhibited an expression pattern similar to Tcm cells, with sustained levels of *CD28*, *CCR7* and *SELL* (L-Selectin/CD62L). Interestingly in both skin-homing and central memory T cell markers, Cluster 11 had consistent increased expression in both phenotype markers compared to the other SS clusters (**Figure 24D**). An additional analysis of gene markers revealed a distinct *FOXP3*$^{+}$ *IL7R*$^{low}$ *TIGIT*$^{+}$ population in cluster 9,

consistent with Treg or Treg-like cells (**Figure 24D**, lower row). All clusters had low and inconsistent expression of *IL2RA* (CD25), another marker of Treg cells, however, consistent with previous reports demonstrating lack of CD25+ Tregs in SS.[337] These data support substantial heterogeneity among the malignant SS cell population and infer a changing transcriptional profile within this clonal population.

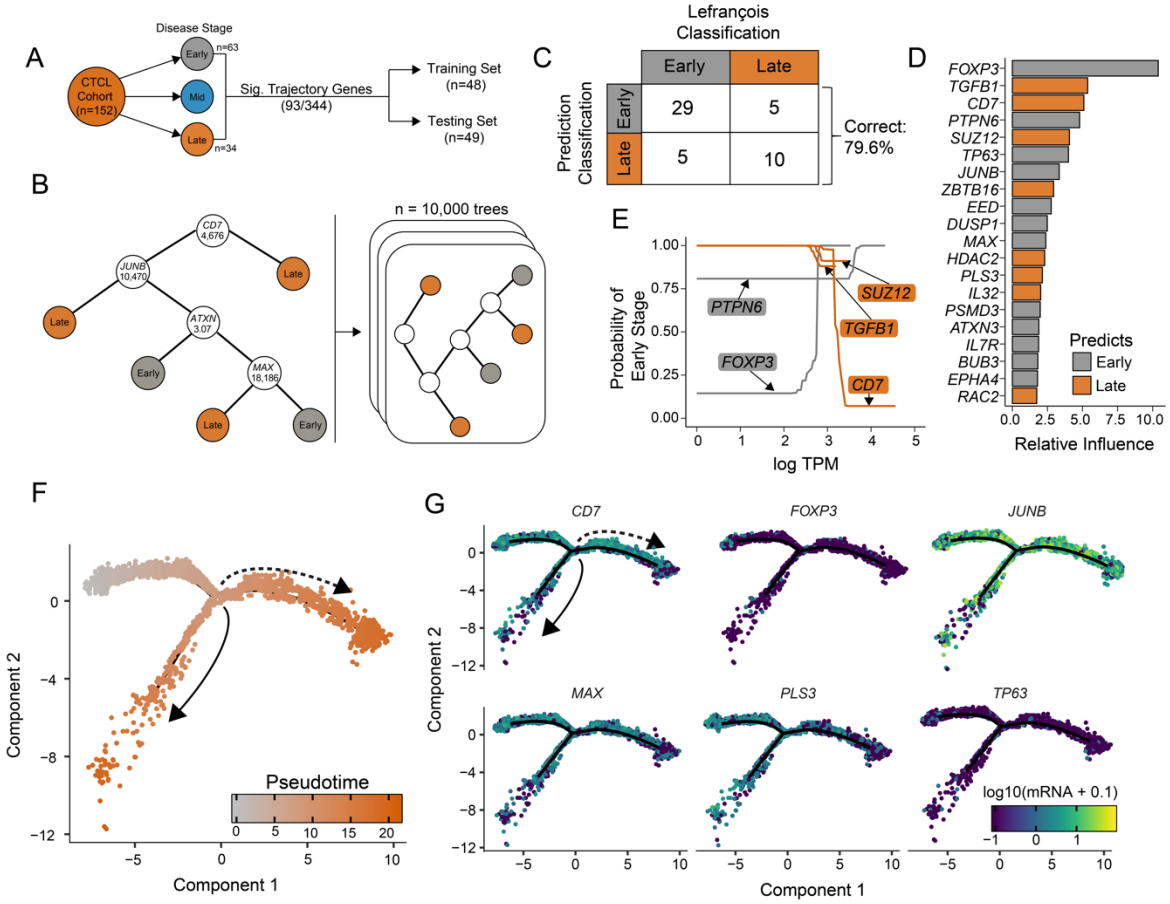


**Figure 25.** Predictive clinical correlates in CTCL using SS single-cell heterogeneity. **A.** Representative schematic of the composition of SRP114956 and the separation into training and testing sets for prediction of clinical stage. **B.** A hypothetical classification decision tree is constructed to predict the CTCL stage based on RNA-sequencing expression data for each patient in the training set (n=48). At each branch in the tree, the patient's transcripts per million (TPM) for a given gene are compared to a cutoff value. If the patient's TPM are below the cutoff, the algorithm proceeds to the left and vice versa, until a terminal classification node is reached. A series of 10,000 boosted trees are grown in sequence utilizing information from previous trees, improving upon previous misclassifications. **C.** The independent test patient data set (n=49) is applied to the 10,000 boosted classification trees and predicted disease states are compared to original classifications. Overall, the boosted decision trees correctly classify 79.6% of the disease states. **D.** The 20 most important genes in generating the boosted classification trees are quantified and displayed in a ranked variable importance plot. Bar color logic is described below. **A.** Partial dependence plots for the five most important variables represent how different levels of gene expression (log TPM) effect the probability of early-disease classification after integrating out the expression of all other genes. Genes with high expression predictive of early disease are colored in grey, while high gene expression more predictive of late stage disease are colored in orange. **F.** Trajectory of malignant cells from clusters 1, 4, 5, 9, and 11 using the Monocle 2 algorithm with pseudo-time represented by the color gradient. **G.** Trajectory and genes expression of selected group of the top 20 genes influencing boosted decision tree.

*Application of artificial intelligence-enabled genetic architecture to single cell SS pseudo-time scheme to predict disease stage*

To better determine and validate if the observed heterogeneity had clinical significance, we downloaded raw sequencing reads from a cohort of CTCL patients.[343] This cohort consisted of 152 CTCL patient samples with targeted sequencing in 344 genes and 3 clinical CTCL classifications: early (Stage ≤ IIA), intermediate/mid (Stages IIB and III), and advanced/late (Stage IV, Figure 4A). In order to improve the separation of the predictions, we isolated early (n=63) and late (n=34) CTCL patients and utilized 93 genes that were predictive of pseudo-time in our single-cell data (**Figure 25A**). After splitting the cohort into training (n=48) and testing (n=49) sets, we constructed a series of boosted classification trees (n=10,000) using the training set (**Figure 25B**). We applied the boosted classification trees to the independent testing set and correctly classified 79.6% of samples into early *versus* late stages (**Figure 25C**). Variable importance was quantified for each gene across the boosted classification trees. The single gene with the largest relative influence in classification was *FOXP3* at 10.39% (**Figure 25D**). Other genes with high relative influence in the classification model include *TGFB1* (5.37%), *CD7* (5.09%), *PTPN6* (4.79%), and *SUZ12* (4.07%). Partial dependence plots for the five most influential genes were constructed to illustrate the effect of each important gene's expression on the probability of early disease stage classification while integrating out other variables (**Figure 25E**). A partial dependence plot was constructed for each of the 20 most important genes (data not shown), and the highest expression level of each gene was compared to the probability of early disease stage classification. Genes with their highest expression predictive of early disease include *FOXP3* and *PTPN6*, while genes with highest expression predictive of late stage disease include *TGFB1*, *CD7*, and *SUZ12*. We next examined the distribution of expression of selected genes from the top 20 genes based on the pseudo-time projection of the manifold (**Figure 25F**,**G**). In this analysis, we noticed inconsistent or irregular expression genes across the cell trajectory manifold when examining the late-stage-CTCL gene predictors (*CD7*, **Figure 25G**). This is possibly a product of the single-cell sequencing of a patient with late-stage SS. In contrast, genes that predict early stage CTCL in the top 20 predictors based on relative influence had consistent decreases in at least one tail of the trajectory manifold (**Figure 25G**). Larger single-cell datasets from patients at different stages of diseases may therefore increase the power of this technique and increase the precision of prognostic and predictive biomarkers.

**Discussion**

Beyond examining transcriptional states of clonal SS cells, this study examines the implications of predicting CTCL progression based on divergent gene drivers. The boosted classification trees demonstrated an efficacious prediction model for classifying early *versus* late disease stage (**Figure 25**). As opposed to the binary evaluation of differential expression of a gene between two different disease states, the boosted classification trees utilizes combinations of continuous expression values associated with early *versus* late disease.[355] Underscoring the value of the boosted classification tree approach was the nearly 80% prediction efficiency for CTCL stage (**Figure 25C**), particularly intriguing considering that feature selection was performed using data

from a single patient. Although limited to a single patient, our analysis provides a framework for the application of single-cell-based high throughput technologies to analyze disease in a clinically meaningful way.

In particular, the expression of *FOXP3* was the most influential predictor of CTCL stage identified from our analysis. FoxP3 is a master transcription factor for regulatory T cells.[356,357] The observation of Treg or Treg-like malignant cells in SS and MF is controversial, with a number conflicting results reported.[336–340] Our work demonstrated decreasing *FOXP3* over purported pseudo-time estimation, and this decrease was associated with an increase in the major Th2 immune driver, *GATA3* (**Figure 24C**). Intriguingly, in the absence of adequate CD25 expression, *bona fide* Tregs retain developmental plasticity, allowing the cells to differentiate into helper T cells dependent on the microenvironment and cytokine milieu.[358] Our data indicate that SS cells may initially express high *FOXP3* and low CD25 (*IL2RA*) and retain similar mutability to FoxP3$^+$CD25$^-$ Tregs.

The maintenance of *FOXP3* expression in Tregs is required to maintain a suppressive phenotype, the loss of which is termed Treg fragility.[359,360] A previous report in SS found a subset of patients with CD25$^-$ FOXP3$^+$ tumor cells, similar to our RNA findings (**Figure 24**), that retain suppressive function.[337] Instead of the malignant proliferation of regulatory T cells first suggested by Berger, *et al*, our data suggests the possibility of a FOXP3$^+$ intermediate state of SS tumor cells.[336] The association with increased *TGFB1* expression with later stage disease, however, would indicate that loss of *FOXP3* does not equate to loss of the ability to elaborate immunoregulatory/suppressive factors. Also of interest, HDAC inhibition, which is effective in treating 30% of SS, has been shown to drive *FOXP3* expression and Treg suppressive function *in vivo*[361] Recent targeted single-cell sequencing of SS cells before and after treatment with HDAC inhibition demonstrated reduction in T cells of the Tcm transcriptional phenotype.[341] Thus, the promotion of an early or intermediate FoxP3$^+$ state in CTCL may be a mechanism of action for vorinostat or other histone deacetylase inhibitors[362] in disease treatment and the prevention of disease progression.

Newer single-cell methods, as used here, allow researchers to characterize and stratify common drivers and sources of heterogeneity in clonal tumors. Similar to our approach, two recent reports in CTCL using transcript-indexed ATAC-seq[363] and limited mRNA sequencing of 110 T-cell-related genes[341] found heterogeneity in malignant SS cells. Beyond the observations of transcriptional heterogeneity, this new level of data provides the opportunity for clinically-meaningful advances in SS and to other cancers. Although limited in scope, our supervised machine-learning approach to predicting CTCL disease state demonstrates the early stages of combining these new highthroughput approaches with predictive algorithms to move beyond simple observations.

# CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

## Conclusions

Despite cancer being singularly ominous as a term, inquiries using molecular biology, biochemistry, and animal modeling can endeavor to slowly resolve fundamental concepts of cancer. These incremental steps forward can lead to improvements in the basic understanding of cancer, as well as translate into meaningful changes in patient care. A force multiplier for these investigations is the proper use of large-scale data in both deductive and inductive reasoning. Starting from the initial work using microarrays by Charles Perou and Therese Sørlie in stratifying breast cancer in the early 2000s[20,21]through moving beyond the TCGA, the field is currently standing at the edge of a fundamentally new ability to quantify tumors. But data without direction is not enough; in turn the field is beginning to harness the data to move beyond observational science and into supervised and deep machine learning to improve both basic and clinical sciences. The latter is the basis of the precision-medicine initiative by the NCI to match therapies with underlying molecular/genetic profiling regardless of the site of origin of the tumors, known as the NCI-MATCH trial.

Although forming 15% of invasive ductal carcinoma diagnoses, BLBC represents an outsized footprint in terms of morbidity and mortality compared to other molecular subtypes.[30] Therapeutic interventions for BLBC have been the same for the last 20 to 30 years, involving surgery, radiation and cytotoxic chemotherapies. With the focus on improving care, we examined protein-level data for over 200 markers and discovered that both MSH2 and MSH6, DNA mismatch repair proteins, were increased in BLBC and were associated with poor overall survival. Although initially investigating DNA mismatch repair as a potential for synthetic lethality in BLBC, similar to the recent successes of PARP inhibtors[36], we found the removal of *Msh2* led to a reduction in tumor growth and could be potentiate response to immune checkpoint blockade. Fitting recent studies into Lynch Syndrome tumors and anti-PD-1 therapies,[153,169] we found the *Msh2* KO lead to increases in T effector cell infiltration and reduction in the suppressive regulatory T cells.

Regulatory T cells in the tumor microenvironment act to dampen anti-tumor immune response. Current therapeutic paradigm for targeting Tregs is the use of anti-CTLA-4 antibodies that are non-specific for tumor-infiltrating Tregs and lead to a high rate of autoimmune complications. Using single-cell RNA sequencing on immune cells from three ccRCC patients, we found distinct transcriptional heterogeneity within tumor-infiltrating Tregs and a novel marker for a suppressive subset of Tregs, *CD177*. CD177$^{+}$ Tregs have increased levels of immune checkpoints, CTLA-4 and PD-1, at the protein level, in addition to other immune suppressive mediators at the mRNA level. Functional analyses of these CD177$^{+}$ Tregs demonstrated increased suppressive ability and increased tumor growth compared to CD177$^{-}$ Tregs. Using the single-cell sequencing expression

profiles, we created a signature and trained a support vector machine using TCGA data, discriminating prognostic groupings in 8 cancer types based on whole-tumor RNA sequencing. The ability to detect transcriptional heterogeneity using single-cell sequencing in tumor-infiltrating immune cells suggests an incomplete picture of the tumor microenvironment. This heterogeneity can also be the key to in identifying better targets for immunotherapies against Tregs with smaller side-effect profiles.

Cutaneous T cell lymphoma, comprised most commonly of mycosis fungoides and Sézary syndrome, is a proliferation of skin-homing $CD4^+$ T cells.[324,325] Using single-cell RNA sequencing of peripheral-blood and malignant $CD4^+$ T cells in a patient with Sézary syndrome, we assessed previously reported and novel markers of the disease. In addition, VDJ sequencing of the individual T cells confirmed the clonal expansion of the Sézary syndrome, with 97.3% of the malignant cells containing the clonotype *TRBV14* (CDR3 amino acid sequence: CASSPLQGTNSPLHF) and *TRAV9-2* (CDR3 amino acid sequence: CALFPNTGFQKLVF) Despite this clonal expansion, we found unique transcriptional states of Sézary syndrome, suggesting the transcriptional evolution from $FOXP3^+$ regulatory-like T cells into T cells that resemble the central memory phenotype. After transcriptional ordering of the Sézary syndrome cells to construct a manifold and pseudo-time estimates, we used the significant genes that change over pseudo-time to perform a boosted tree classification system to estimate cutaneous T cell lymphoma disease state using a secondary cohort of 152 patients who received targeted sequencing of 344 genes, predicting the state 79.6% of the time.

**Future directions**

The goal of this dissertation is to develop translational strategies for prognosis and therapies using large-scale data. It serves, at least in a small part, as a proof-of-concept for the use of big data and bioinformatics to frame hypotheses in real time. Part of the nature of having a collection of diverse topics that combine to form a dissertation is the number of questions left unanswered. But to me, that is the best part and future directions by chapter order include:

*Chapter 2*

Investigation into the underlying mechanism of the activation of anti-tumor immunity elicited by the knockout of mismatch DNA repair. Our preliminary data suggests a tumor-burden independent mechanism by which the removal of DNA mismatch repair leads to increases in antigen presentation and IFN-$\gamma$ secretion/responsiveness. Teasing apart the type 1 and type 2 interferon contribution from tumor-infiltrating immune cells *versus* tumor cells will be the next vital step to understand the underpinning mechanism. Beyond the scope of this project, we have also been screening inhibitors of DNA mismatch repair as a therapeutic option for the aggressive BLBC. In addition, we have recently bred *Msh2*$^{flux/flux}$ mice into the C3Tag mouse

model of basal-like breast cancer[364] with epithelial compartment-specific expression of Cre, allowing for the further investigation of MSH2 as a novel tumor promoter for BLBC in a spontaneous tumor model.

*Chapter 3*

The heterogeneity of tumor-infiltrating Tregs provides a starting point for the development of more specific immunotherapies for Tregs in the tumor microenvironment. Our initial data suggests that CD177 may be a specific marker of suppressive tumor-infiltrating Tregs. To this end, we are in the process of developing *Cd177*$^{flux/flux}$ mice to remove *Cd177* in a Treg-specific manner to investigate the modulation of tumor burden and the functional role of *Cd177* in Tregs. Furthermore, preclinical inquiry into the potential of therapeutically targeting CD177 using *Cd177*$^{flux/flux}$ mice as controls.

*Chapter 4*

The development and progression of Sézary syndrome is not well understood and has been thought to be, in part, an immune reaction to antigen presentation of Langerhans cells in the skin. Using single-cell RNA sequencing of skin lesions and malignant blood cells may offer better insights to the initiation and progression Sézary syndrome. Additionally, the sequencing of the dendritic-like Langerhans cells within skin lesions, at the potential interface of immune stimulation of malignant Sézary syndrome cells, may offer insights into new potential therapies.

# APPENDIX

## REEVALUATING E-CADHERIN AND β-CATENIN: A PAN-CANCER PROTEOMIC APPROACH WITH EMPHASIS IN BREAST CANCER

### Rationale

E-Cadherin is a structural component of adherens junctions, linking the actin cytoskeleton to adjacent cells forming epithelial tissues.[365] Throughout carcinomas (cancers of epithelial origin), E-Cadherin expression has been inversely correlated with tumor stage, pathological stage, and prognosis.[366–371] Intuitively, the loss of E-Cadherin in carcinomas is thought to encourage invasion and metastasis via loss of cell-to-cell interactions.[372,373] The loss of E-Cadherin has been used as a hallmark for epithelial-mesenchymal transition (EMT) genetic reprogramming that changes the epithelial characteristics of cancer cells.[48] However, recent studies have called into question whether cancer cells require the loss of E-Cadherin or EMT to invade and metastasize.[11-13] Moreover, several studies have suggested the clustering of tumor cells via adherens or adherens-like junctions may facilitate metastasis.[374]

Catenins, including α-Catenin, β-Catenin, and p120-Catenin, are intracellular components of the adherens junction. Catenins provide structural support as part of the adaptor complex that attaches the actin cytoskeleton to E-Cadherin. Additionally, β-Catenin is a transcriptional co-activator in the WNT signaling pathway with established roles in embryogenesis, stem cell regulation, carcinogenesis, and EMT.[375] E-Cadherin is believed to sequester β-Catenin to the plasma membrane at a one-to-one ratio.[376,377] With the presence of E-Cadherin, structural β-Catenin is prevented from participating in WNT ligand-mediated signaling.[378] However, it is unknown whether deficiency in E-Cadherin is sufficient to drive β-Catenin activation as the loss of E-Cadherin is often accompanied by co-loss of β-Catenin in breast cancer.[379–381]

Breast cancer is one of the few cancer types for which E-Cadherin and β-Catenin have been investigated for diagnostic, prognostic, and mechanistic value with inconsistent and sometimes contrasting conclusions.[371,382–385] Invasive breast cancer is divided into two major subgroups based on histologic and molecular traits, including infiltrating ductal carcinoma (IDC) with 70-80% of total incidence and infiltrating lobular carcinoma (ILC) with 10-15% of total incidence.[386] Clear distinction in prognosis based on ILC *versus* IDC has been disputed, with groups finding no difference,[387] ILC having better prognosis,[388–390] and ILC having worse prognosis.[389,391] Recent investigation by the Cancer Genome Atlas (TCGA) network has demonstrated distinct molecular subtypes in ILC that differ in overall and disease-free survival, which may account for a portion of the mixed results.[392]

A defining feature of many ILC lesions is the loss of E-Cadherin staining via IHC. This loss of E-Cadherin is a result of truncating mutations in the E-Cadherin gene, *CDH1*, seen in up to 50% of ILC or epigenetic silencing of *CDH1* in up to 41% of ILC.[373,393,394] Concomitant with *CDH1* mutations is the loss of 16q, the location of *CDH1,* and is thought to be an early event in the development of ILC.[392] In contrast, loss of E-Cadherin in IDC has been considered a precursor step to invasion and metastasis, with controversial associations with both higher grade and pathological stage.[371,384,395,396] Interestingly, IDC lymph node and distant metastases can be positive for E-Cadherin staining, suggesting a re-expression of E-Cadherin.[371,397] While the literature has indicated a poor prognostic association of reduced or loss of E-Cadherin in IDC, a number of reports have found no association between E-Cadherin status and tumor stage, lymph node status, presence of metastatic lesion, or recurrence-free survival.[382–384] In a nuclear-grade controlled breast cancer cohort of 470 specimens, the expression of E-Cadherin and Catenins were directly associated with shorter patient survival, casting doubt on the commonly believed good prognostic indication of E-Cadherin and Catenins in breast cancer.[385]

As a central node between cell adhesion and stem-like/mesenchymal signaling, the adherens junction has been investigated for both prognosis and mechanistic underpinnings in carcinomas. Here, we reexamined the relationship between E-Cadherin and β-Catenin in 19 cancer types, with a focus on their prognostic values using computational and immunohistochemical approaches. We found a general agreement of E-Cadherin and β-Catenin as good prognosis markers in majority of cancers, with breast cancer and kidney papillary cell carcinoma as outliers.

**Materials and methods**

*TCGA data analysis*

Publicly available replicate-based normalized (RBN) reverse-phase protein array (RPPA) data from the Cancer TCGA Pan-Cancer Analysis datasets was downloaded from the MD Anderson TCPA.[398] Clinical data matrices were downloaded from the UCSC Cancer Browser.[399] A total of 5,144 samples contained both RPPA and clinical information and were subdivided by cancer type. RBN protein levels of E-Cadherin and β-Catenin were compared across the Pan-Cancer datasets using Z-score scaled protein level. Survival analysis was performed using R packages: survival (v2.41-3), survMisc (v0.5.4), and survminer (v0.4.0). Graphical display of the hazard ratio summary across the Pan-Cancer cohorts utilized –log10(P-value) and Cox proportional hazard ratio with size of points based on the –log10(P-value). Survival analysis for *CDH1* utilized the optimal-cutoff function available through survMisc for the TCGA data, Kaplan-Meier Plotter cohort,[400] and the METABRIC patient cohort.[401] The heatmap for the invasive breast carcinoma (BRCA) cohort of the TCGA dataset was made using the heatmap.3 function in R. β-Catenin to E-Cadherin ratio was calculated as a quotient of the two factors with mean-centralization of natural-log transformation.

*Tumor samples*

Formalin-fixed, paraffin-embedded cancer samples from breast cancer excisions were obtained from the surgical pathology archives of the University of Iowa Hospitals and Clinics. Samples were derived from two cohorts: 1) consecutive cases of excised $ER^-$ breast cancer with sufficient tissue to put in TMA and 2) a smaller group of consecutive excised $ER^+$ breast cancer. Pathologic data recorded from patient pathology reports included the following: tumor size, T stage, lymph node status, and ER/PR/HER2 biomarker status. Survival data was available for 30 of the 163 patients. ER and PR (IHC) and HER2 (IHC and/or fluorescence *in situ* hybridization) status had been assessed under the context of patient care.

*TMA construction and IHC*

Tissue microarrays (TMA) were constructed using the Manual Tissue Arrayer MTA-1 (Beecher Instruments; Sun Prairie, WI) with tumors arrayed in triplicate 1 mm cores. IHC was performed on 4-μm tissue sections on a Dako Autostainer Link 48 (Carpinteria, CA) after deparaffinization, rehydration, and heat-induced epitope retrieval with Tris/EDTA Target Retrieval Solution (pH 9) on the Dako PT Link. Mouse monoclonal antibodies to β-Catenin (clone β-Catenin-1; 1:500; 15-minute primary and secondary antibody incubations; Dako) and E-Cadherin (clone NCH-38; 1:50; 15-minute primary and secondary antibody incubations; Dako) were used. Specific signal was visualized with the polymer-based Dako EnVision Flex System, with 3'-diaminobenzidine as the chromogen. Immunostained slides were counterstained with Harris hematoxylin (Leica Biosystems; Buffalo Grove, IL) and coverslipped. The positive control tissues for β-Catenin and E-Cadherin were solid pseudopapillary neoplasm and breast cancer, respectively.

*Immunohistochemical scoring*

IHC was scored by two pathologists at a double-headed light microscope. For β-Catenin and E-Cadherin, one of the following three patterns was assigned: intact membranous (crisp, complete staining readily observable at 20-40x magnification), reduced membranous (weaker, incomplete staining), and absent (complete absence of cell membrane staining). β-Catenin was additionally assessed for nuclear staining.

*Statistical analysis and visualizations*

Hazard ratios and *P*-values were calculated using the Cox proportional hazard function in R comparing the upper *versus* lower quartile of the indicated proteins. Significance testing for sampling distribution was conducted using the $\chi^2$ test. Correlations were calculated with the psych R package (v1.7.8) with the Spearman rank-based approach, which is more robust for outliers. Analyses and graphical plotting were performed using RStudio and the ggplot2 R package (v2.2.1) and the ggridges (v0.4.1) for the density-based histogram plots or ridge plots. Graphical summary of the data sources, data analysis, and results is provided in **Figure 26**.

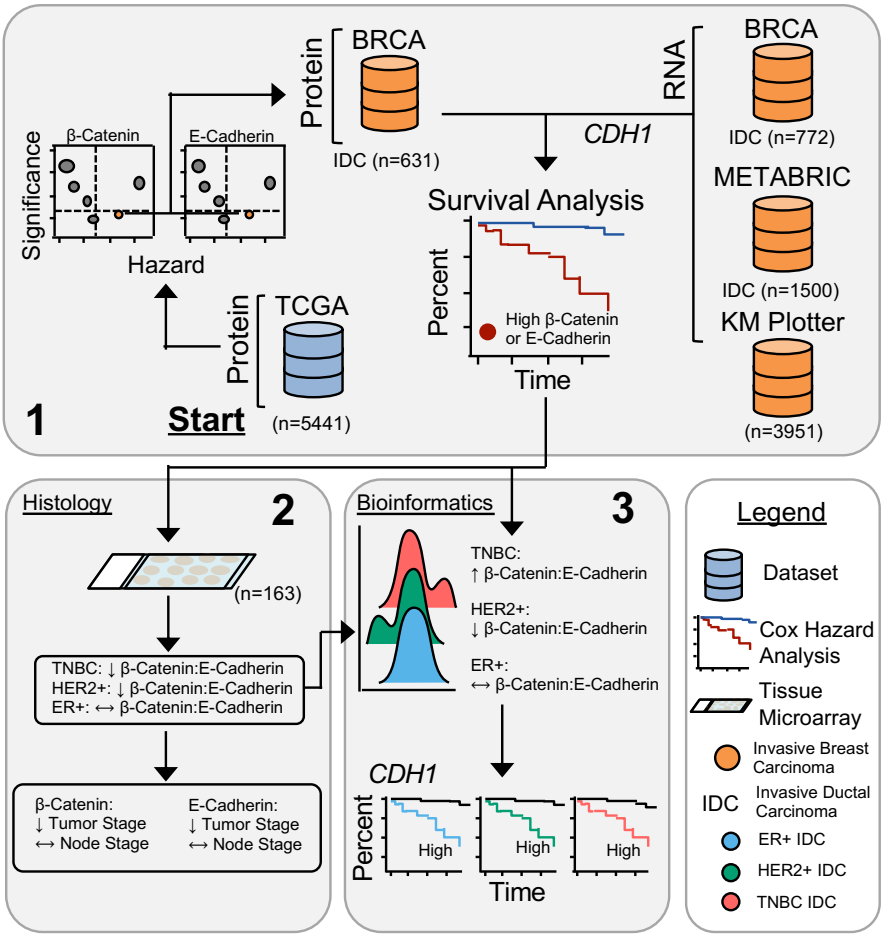


**Figure a.1.** Graphical summary of results and workflow of the analysis. **1.** starting with survival analysis of β-catenin and E-cadherin across 5,441 samples in the Pan-Cancer the Cancer Genome Atlas (TCGA). Further focus on breast invasive cancer (BRCA) found increased β-catenin and E-cadherin as poor prognostic indicators in infiltrating ductal carcinoma (IDC), mirroring survival analysis of *CDH1* in 6,465 patient samples. This led to the further investigation of β-catenin and E-cadherin staining and localization using tissue microarrays (**2**) and further characterization of β-catenin *versus* E-cadherin dynamics and survival using bioinformatics **3.** ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; KM, Kaplan-Meier; TNBC, triple-negative breast cancer.

## Results

*β-Catenin and E-Cadherin protein correlation and predictive ability across cancers*

With the recent availability of the reverse-phase protein array (RPPA) data for the TCGA samples developed by MD Anderson, we examined the protein level of β-Catenin and E-Cadherin in 19 cancer types (**Figure 27A**). Across these TCGA cancer cohorts, we found a consistent positive correlation between β-Catenin and E-Cadherin protein level, with Spearman ρ greater than 0.5 in 15 of 19 cancer types. Low correlations between β-Catenin and E-Cadherin were observed in adrenocortical carcinomas (ACC, ρ=-0.321, *P*=0.0297), glioblastoma multiforme (GBM, ρ=0.199, *P*=0.00470), kidney clear cell carcinoma (KIRC, ρ=0.275, *P*=3.97e-09), and low-grade glioma (LGG, ρ=-0.144, *P*=0.0213). Using the Cox regression analysis, we compared the highest *versus* lowest quartile of samples in each cancer type for β-Catenin and E-Cadherin. β-Catenin exhibited a range of hazard ratios, either as a good or as a poor prognostic indicator dependent on the cancer type (**Figure 27B**). However, with a significance threshold of *P*≤0.05, only three cancer types with β-Catenin as a good prognostic indicator met the criteria: KIRC (HR=0.688), head and neck squamous cell carcinoma (HNSC, HR=0.786) and GBM (HR=0.853). In contrast, high E-Cadherin was correlated with good prognosis in 12 of

the 19 cancer types, with two cancer cohorts reaching significance threshold of $P \leq 0.05$ (**Figure 27C**). Notably E-Cadherin was a poor prognosis marker in kidney papillary cell carcinoma (KIRP, HR=1.537, $P=0.027$) and in the breast invasive carcinoma (BRCA, HR=1.127, $P=0.2$) cohorts of TCGA datasets (**Figure 27C**). Despite being below the significance threshold of $P \leq 0.05$, we focused on breast cancer as it is a large cohort of patient specimens with a variety of histological and subtypic designations that could be obscuring the prognostic prediction of E-Cadherin or β-Catenin.

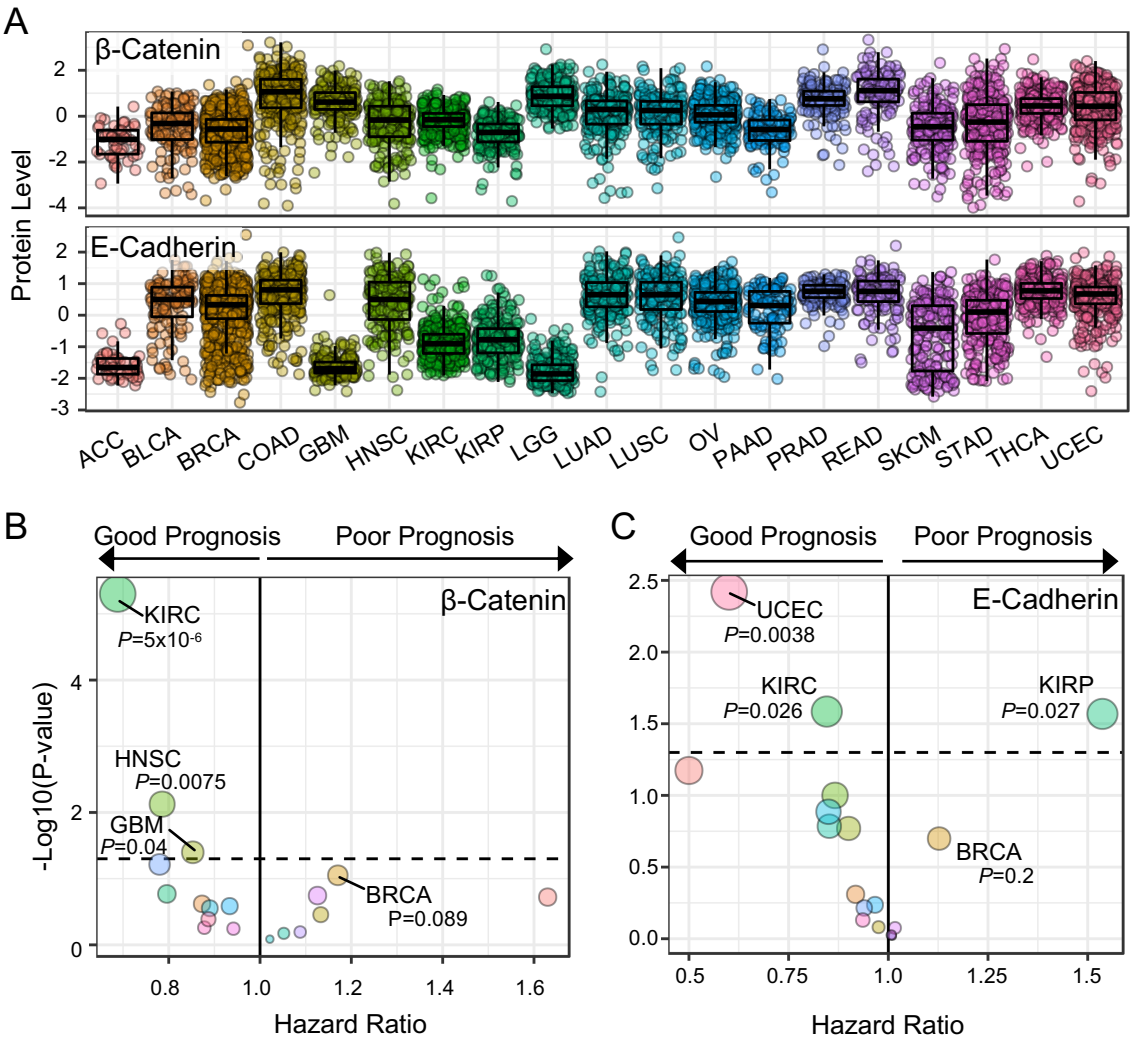


**Figure a.2**. Protein level and prognostic value of β-Catenin and E-Cadherin across 19 cancer types. **A**. Z-score protein level of β-Catenin (top) and E-Cadherin (bottom) in the TCGA Pan-Can cohort ($P<0.001$ for β-Catenin and E-Cadherin). **B**. Summary of Cox proportional hazard regression comparing the upper quartile *versus* lowest quartile of β-Catenin protein level. Dotted line indicates a $P=0.05$. **C**. Summary of Cox proportional hazard regression comparing the upper quartile *versus* lowest quartile of E-Cadherin protein level. Dotted line indicates a $P = 0.05$. Size of points is on a relative scale based on –log10(*P*-value). ACC, adrenocortical carcinoma (n=46); BLCA, bladder urothelial carcinoma (n=127); BRCA, breast invasive carcinoma (n=820); COAD, colon adenocarcinoma (n=326); GBM, glioblastoma multiforme (n=201); HNSC, head and neck squamous cell carcinoma (n=203); KIRC, kidney clear cell carcinoma (n=444), KIPR, kidney papillary cell carcinoma (n=207); LGG, lower grade glioma (n=257); LUAD, lung adenocarcinoma (n=233); LUSC, lung squamous cell carcinoma (n=192); OV, ovarian serous cystadenocarcinoma (n=408); PAAD, pancreatic adenocarcinoma (n=105); PRAD, prostate adenocarcinoma (n=164); READ, rectum adenocarcinoma (n=127); SKCM, skin cutaneous melanoma (n=207); STAD, stomach adenocarcinoma (n=299); THCA, thyroid carcinoma (n=374); UCEC uterine corpus endometrial carcinoma (n=404).

*β-Catenin and E-Cadherin protein in histological subtypes of breast cancer*

Previous literature has reported the loss of E-Cadherin as a diagnostic marker of ILC, together with loss of catenins.[373,393,394,402] Using unsupervised clustering of β-Catenin, E-Cadherin, and other components involved in the maintenance or signaling of the adherens junction, we found the ILC (purple) clustered distinctly compared to other histological diagnoses (**Figure 28A**). The ILC cluster was defined by significant decreases in β-Catenin (**Figure 28A**,**B**) and E-Cadherin (**Figure 28A**,**C**). Conversely, IDC varied in β-Catenin and E-Cadherin protein levels and were significantly correlated with each other (Spearman ρ=0.64, *P*<1e-15 ). In IDC, E-Cadherin was inversely correlated with N-Cadherin, a marker of EMT, (ρ=-0.340, *P*=1.33e-15) compared to ILC (ρ=0.301, *P*=0.0108). E-Cadherin-to-N-Cadherin switching has been previously seen as a stage- and grade-dependent phenomenon in IDC.[403] After subsetting the IDC patient samples, Cox regression analysis demonstrated β-Catenin (**Figure 28D**) and E-Cadherin (**Figure 28E**) proteins as poor prognostic indicators. In contrast, neither β-Catenin (**Figure 2F**, HR=1.126, *P*=0.67) nor E-Cadherin (**Figure 2G**, HR=1.149, *P*=0.51) was predictive for prognosis in ILC patients.

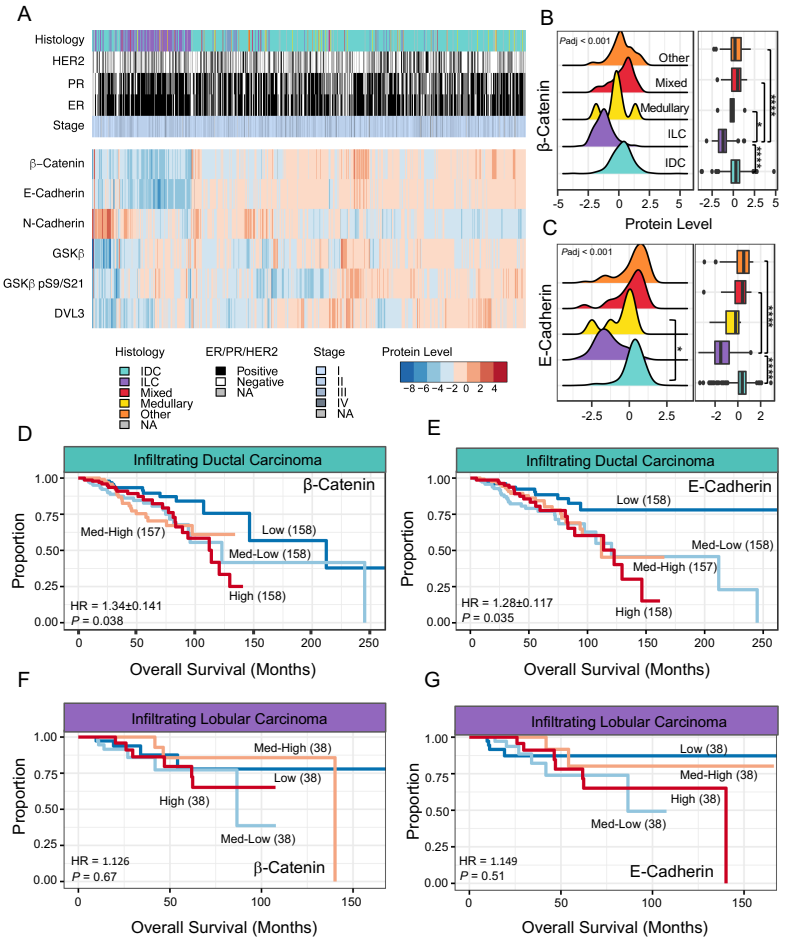


**Figure a.3**. The dynamics of β-Catenin and E-Cadherin in breast invasive carcinoma. **A**. Heatmap of the clinical and pathologic features of TCGA BRCA cohort (n=873). Samples across the BRCA cohort underwent unsupervised clustering based on the indicated protein levels. **B.** Density-based histogram and corresponding boxplot of β-Catenin protein level by histologic diagnosis. IDC, Infiltrating Ductal Carcinoma (n=631); ILC, Infiltrating Lobular Carcinoma (n=152), Medullary Carcinoma (n=6), Mixed Histology (n=23), Other (n=60). Adjusted *P*<0.001. **C**. Density-based histogram and corresponding boxplot of E-Cadherin protein level by histologic diagnosis. Adjusted *P*<0.001. **D**, **E** Kaplan-Meier curve using β-Catenin (**D**) and E-Cadherin (**E**) protein level to split IDC samples into quartiles, using Cox proportional hazard regression comparing the upper quartile *versus* lowest quartile. **F**, **G** Kaplan-Meier curve using β-Catenin (**F**) and E-Cadherin (**G**) protein level to split ILC samples into quartiles, using Cox proportional hazard regression comparing the upper quartile *versus* lowest quartile. *$P$<0.05,****$P$<0.0001. ER, estrogen receptor; progesterone receptor; HER2, human epidermal growth factor receptor 2.

To further corroborate the finding of E-Cadherin as a poor prognostic indicator in IDC, we queried mRNA-based datasets. In the TCGA BRCA cohort, we first examined the correlation between *CDH1* mRNA and E-Cadherin and found the Spearman ρ=0.53 (**Figure 29A**), indicating a significant positive correlation between RNA and protein levels. Similar to the protein survival analysis, *CDH1* mRNA expression was a poor prognosis indicator across all IDC in the TCGA BRCA cohort (**Figure 29B**, N=772, HR=1.94). We observed the same trend in a second (**Figure 29C**, N=3,951, HR=1.37, KM Plotter) and a third (**Figure 29D**, left panel, N=1,500, HR=1.32) breast cancer cohort consisting of normalized microarray data.[401,404] Similar to the protein data, *CDH1* mRNA was not predictive of prognosis in ILC patients (**Figure 29D**, right panel), which may be complicated by low mRNA and protein expression in ILC. There was no significant correlation between *CTNNB1* mRNA and β-Catenin protein (ρ=0.092, data not shown), discounting the validity of using mRNA as a predicative marker for breast cancer.

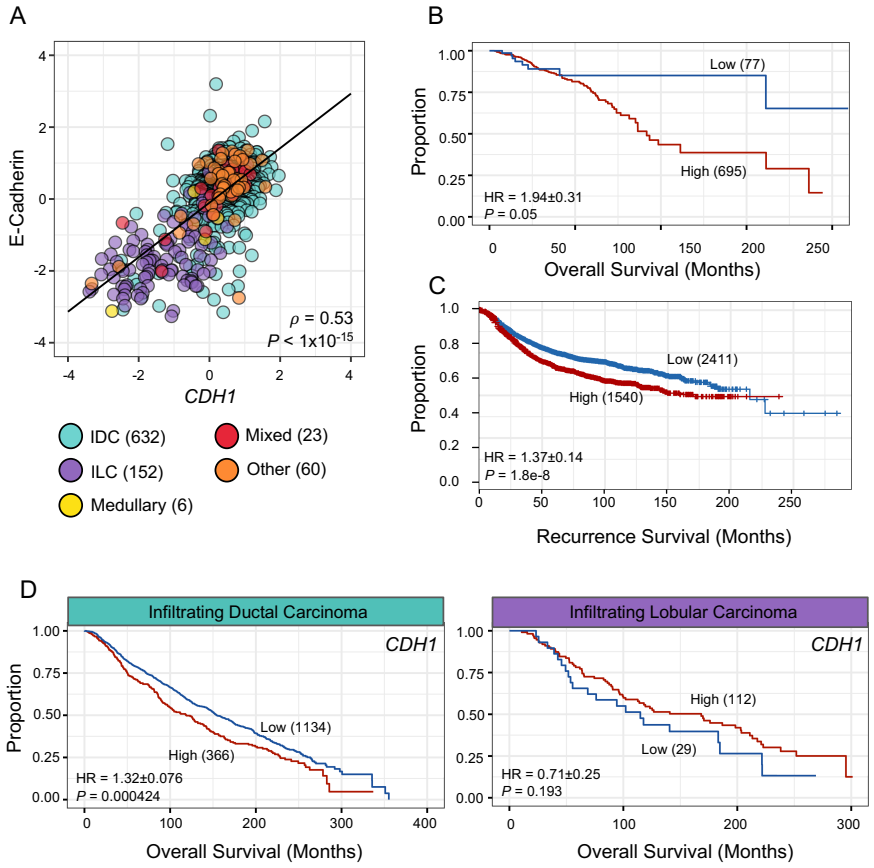


**Figure a.4.** mRNA comparisons of *CDH1* in breast cancer. **A.** Scatterplot of E-cadherin and *CDH1* across the invasive breast carcinoma (BRCA) The Cancer Genome Atlas cohort. **B**. Kaplan-Meier curve from the BRCA cohort using *CDH1* levels split by auto-cutoff for high *versus* low samples, using Cox proportional hazard regression. **C.** Kaplan-Meier curve from the Kaplan-Meier Plotter cohort using *CDH1* levels split by auto-cutoff for high *versus* low samples, using Cox proportional hazard regression. **D.** Kaplan-Meier curve from the METABRIC cohort using *CDH1* levels split by auto-cutoff for high *versus* low samples, using Cox proportional hazard regression and separated by histologic diagnosis of either infiltrating ductal carcinoma (IDC; left panel) or infiltrating lobular carcinoma (ILC; right panel).

*Immunobiological evaluation of β-Catenin and E-Cadherin in IDC*

The poor prognostic indication of E-Cadherin in the IDC samples led us to examine β-Catenin and E-Cadherin status using IDC patient specimens from the University of Iowa Hospital and Clinics. We constructed tissue arrays including 220 specimens with known clinical information. Of the initial cohort, 163

**Table 2.** Summary of β-Catenin and E-Cadherin IHC tissue microarray.

| *Parameter* | *Cohort* | *Total* |
|---|---|---|
| *Tumor Stage* | | |
| Tis (DCIS)* | 4 | |
| T1 | 73 | 163 |
| T2 | 61 | |
| T3 | 9 | |
| T4 | 9 | |
| NA† | 7 | |
| *Node Stage* | | |
| N0 | 68 | |
| N1 | 37 | |
| N2 | 15 | 163 |
| N3 | 17 | |
| NX | 11 | |
| NA† | 15 | |
| *Biomarker* | | |
| ER+ | 16 | |
| HER2+ | 28 | 108 |
| TNBC | 64 | |
| *Histological Diagnosis* | | |
| DCIS | 9 | |
| IDC | 137 | |
| ILC | 7 | |
| Medullary | 2 | 163 |
| Metaplastic | 1 | |
| Mucinous | 3 | |
| Papillary | 1 | |
| NA† | 3 | |

***** Ductal carcinoma in situ (Tis/DCIS)
† Information not available (NA)

of 220 cases were further analyzed for IHC staining. This reduction from 220 to 163 was due to incomplete clinical information or fragment and absent cores in the tissue microarray (**Table 2**). Samples were scored based on the localization of β-Catenin or E-Cadherin and the strength of the IHC staining (**Figure 30**). Of the 163 cases, there were four staining patterns for β-Catenin: 47 samples showed membranous β-Catenin (**Figure 30A**), 94 had a reduced pattern (**Figure 30C**), 20 were absent (**Figure 30E**), and 2 samples displayed a component of nuclear β-Catenin staining (**Figure 30G**). Both samples with nuclear staining for β-Catenin had corresponding strong membranous E-Cadherin staining (**Figure 30H**). In contrast, there were three patterns of E-Cadherin staining: 81 cases had an intact membranous E-Cadherin (**Figure 30B**), 71 reduced E-Cadherin (**Figure 30D**), and 11 lacked staining (**Figure 30F**).

*β-Catenin and E-Cadherin status in IDC by clinical subtype*

We assessed β-Catenin and E-Cadherin staining patterns according to breast cancer biomarker status or clinical subtype, *i.e.*, ER+, ER-/HER2+ (referred to herein as HER2+), or triple-negative breast cancer (TNBC). Of the 163 samples analyzed, 108 samples could be assigned as ER+, HER2+, or TNBC (Table 2). The E-Cadherin IHC expression patterns showed higher proportions of intact membranous E-Cadherin staining in ER+ tumors and a high percentage of reduced IHC expression in TNBC ($P = 0.0011$). Additionally, HER2+ tumors had intact staining pattern of E-cadherin. The same significant association was not seen with β-Catenin and biomarker status ($P = 0.439$). Notably, HER2+ samples had the least pronounced correlation between β-Catenin and E-Cadherin, with 78.5% of samples with intact membranous E-Cadherin and only 25% of samples with membranous β-Catenin staining (**Table 3**).

**Table 3.** Summary of IHC staining for E-Cadherin and β-Catenin by subtype. Table based on 108 of the 163 samples designated as TNBC, triple-negative, ER+, or HER2+.

| *Diagnosis* | *Samples* | *E-Cadherin* | | | *β-Catenin* | | |
|---|---|---|---|---|---|---|---|
| | | *Membranous* | *Reduced* | *Absent* | *Membranous* | *Reduced* | *Absent* |
| *TNBC* | 64 | 24 | 33 | 5 | 18 | 42 | 5 |
| *ER+* | 16 | 10 | 5 | 1 | 6 | 7 | 3 |
| *HER2+* | 28 | 22 | 5 | 1 | 7 | 18 | 3 |
| | | *P*=0.011 | | | *P*=0.439 | | |

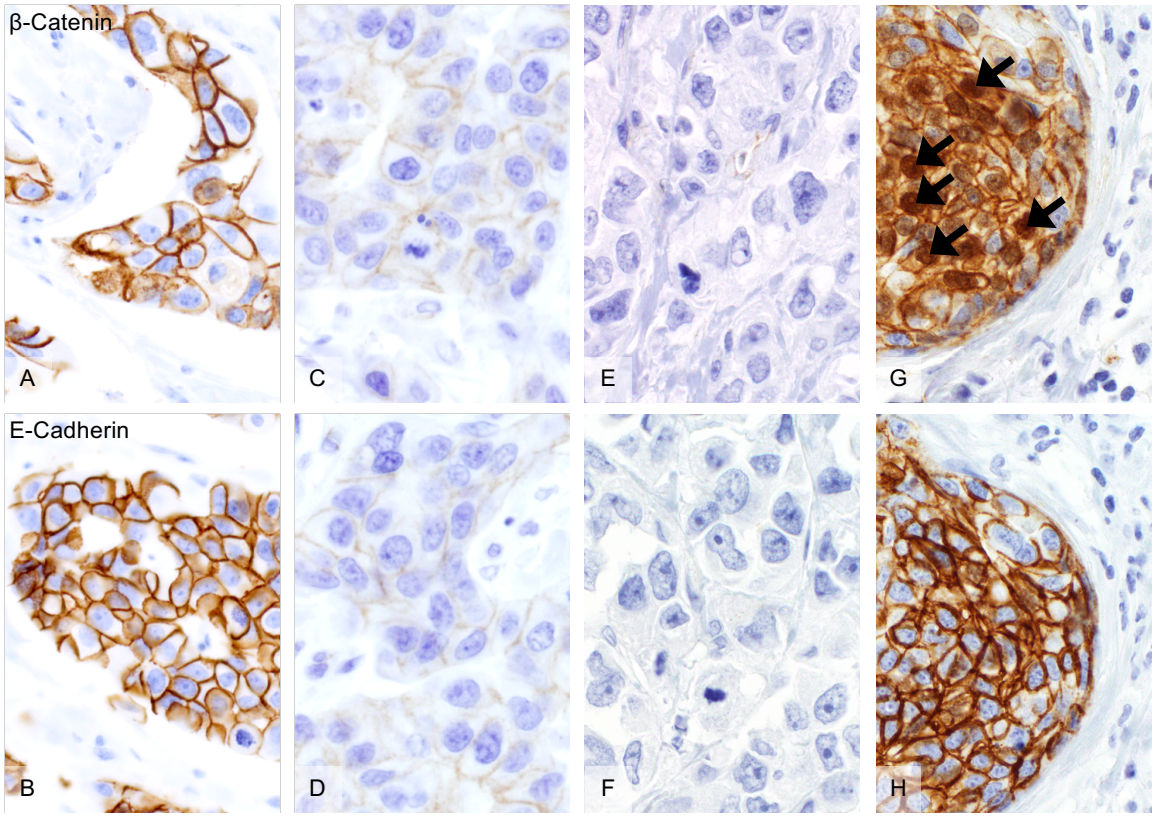


**Figure a.5.** IHC staining results for β-Catenin and E-Cadherin in breast cancer samples. Representative IHC staining patterns that were observed: **A**,**B**. membranous staining for both β-Catenin (top) and E-Cadherin (bottom). **C**,**D**. reduced membranous expression for both β-Catenin (top) and E-Cadherin (bottom). **E**,**F**. absent staining for both β-Catenin (top) and E-Cadherin (bottom). **G**,**H**. membranous and nuclear staining of β-Catenin with strong E-Cadherin membranous staining (arrows indicate cells with nuclear staining of β-Catenin). Images were taken at 400x magnification.

Comparing our IHC results, we used the RPPA data for IDC samples and classified them into clinical subtypes based on IHC confirmation of $ER^+$ (n=358), $HER^+$ (n=29) or TNBC (n=63). In order to examine the relationship of β-Catenin and E-Cadherin in the context of the adherens junction, we first examined the correlation between the two proteins across IDC samples (**Figure 31A**, $\rho$=0.64). Across IDC, 22 samples were designated as outliers (**Figure 31A**, dotted circle) with E-Cadherin protein level < -3 and a β-Catenin > -3. There was no significant difference in these samples by biomarker, tumor stage, nodal stage, or survival compared to the rest of the IDC cohort. Using the mean-centralized ratio of β-Catenin to E-Cadherin, a method of identifying outliers in either β-Catenin or E-Cadherin protein level, we produced density-based histograms to examine subgroup-specific differences in the relationship (**Figure 31B**). Similar to our IHC results, $HER2^+$ samples (green) had 8 of 29 samples with β-Catenin-to-E-Cadherin ratio below the identified cut-off, indicating a relative decrease in β-Catenin in $HER2^+$ samples. In contrast, $ER^+$ patient samples (blue) had a normal distribution of β-Catenin-to-E-Cadherin ratio. TNBC samples (red) had a relative increase in β-Catenin-to-E-Cadherin ratio that is inconsistent with the IHC results that found a greater reduction in β-Catenin compared to E-Cadherin in TNBC (**Table 3**). However, this higher level of β-Catenin is consistent with literature that

suggests β-Catenin can drive TNBC and is a poor prognostic indicator for TNBC patients.[405] We were unable to perform survival analysis using the TCGA BRCA cohort due to the limited size and high degree of censored samples in the HER2/TNBC subgroups. Alternatively, we used the METABRIC IDC samples, separating the samples by IHC-confirmed clinical subtype. Despite the difference in β-Catenin-to-E-Cadherin dynamics, *CDH1* was a poor prognostic indicator in ER+ and HER2+ IDC (**Figure 31C**). The highest *CDH1* levels in TNBC was associated with poor prognosis with a *P*-value slightly higher than the $P<0.05$ cut-off (**Figure 31C**, right panel).

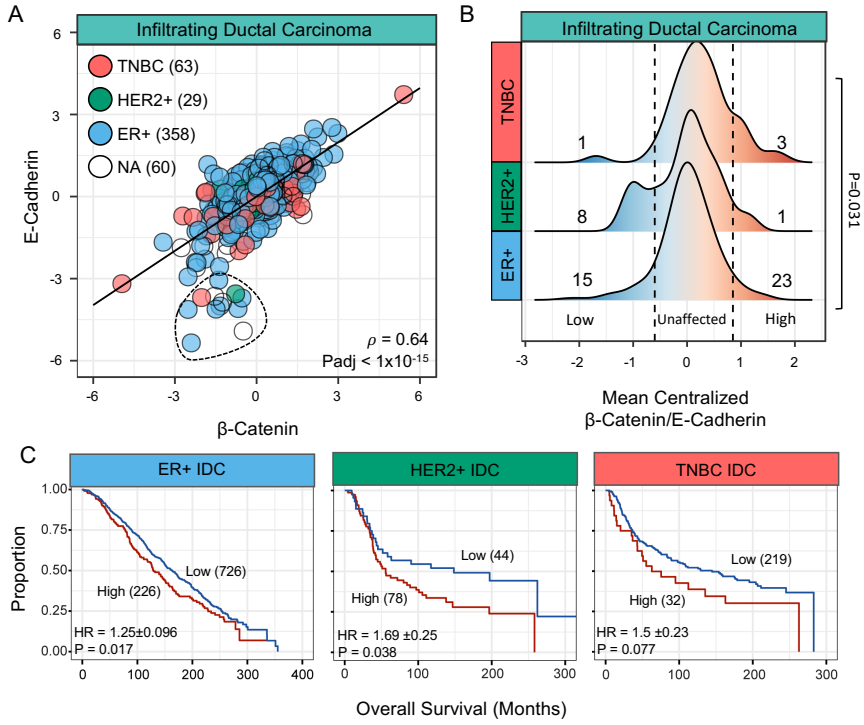


**Figure a.6.** β-Catenin and E-Cadherin protein in IDC samples by subtype. **A**. Scatterplot of β-Catenin and E-Cadherin protein across IDC samples. Outliers, defined as E-Cadherin < -3 and β-Catenin > -3 circled with dotted line. **B**. Density histogram of mean-centralized quotient of β-Catenin and E-Cadherin. Cut-off value of -0.65 (low, left-bound dotted line) and 0.8 (high, right-bound dotted line) were used based on distinct subpopulations. $\chi^2$ tests were performed comparing low, high, and unaffected population distributions. **C.** Kaplan-Meier curve from the ER+ (blue), HER2+ (green), and TNBC (red) samples in the METABRIC cohort using *CDH1* levels split by auto-cutoff for high *versus* low samples, using Cox proportional hazard regression.

*β-Catenin and E-Cadherin status in IDC by stage*

Using our IHC staining, we next investigated the trend in β-Catenin and E-Cadherin staining patterns by the American Joint Committee on Cancer (AJCC) tumor and node stage (**Table 4**). Both β-Catenin and E-Cadherin demonstrated statistically significant negative correlation with tumor stage. In late stage tumors (T2 through T4), β-Catenin and E-Cadherin were significantly reduced or absent compared with the T0 and T1 stages. β-Catenin and E-Cadherin status was less predictive of lymph node involvement (**Table 4**). However, the characteristics between the sample distribution of β-Catenin and E-Cadherin by lymph node status varied. E-Cadherin had relative equal sample distribution between each node status with reduced E-Cadherin seen in 40-45.5% of samples in each node stage. Conversely, β-Catenin had bimodal distribution of reduced staining with peaks at N0 (62.3%) and N2-3 stages (80%).

**Table 4.** Summary of IHC staining for β-Catenin and E-Cadherin by tumor and nodal stage.

| *Target* | *Parameter* | *n* | *Membranous* | *Reduced* | *Absent* | *P-value* |
|---|---|---|---|---|---|---|
| *β-Catenin* | *Tumor Stage* | 144 | | | | |
| | T0 | | 2 | 1 | 1 | |
| | T1 | | 26 | 36 | 5 | *P*=0.0258 |
| | T2 | | 11 | 37 | 7 | |
| | T3+T4 | | 3 | 12 | 3 | |
| | *Nodal Status* | 133 | | | | |
| | N0 | | 21 | 43 | *5* | |
| | N1 | | 11 | 17 | *5* | *P*=0.0834 |
| | N2 | | 3 | 12 | *0* | |
| | N3 | | 4 | 12 | *0* | |
| *E-Cadherin* | *Tumor Stage* | 153 | | | | |
| | T0 | | 3 | 0 | 0 | |
| | T1 | | 44 | 26 | 3 | *P*=0.00541 |
| | T2 | | 22 | 33 | 3 | |
| | T3+T4 | | 6 | 8 | 5 | |
| | *Nodal Status* | 135 | | | | |
| | N0 | | 33 | 31 | 4 | |
| | N1 | | 17 | 16 | 2 | *P*=0.955 |
| | N2 | | 8 | 6 | 1 | |
| | N3 | | 8 | 8 | 1 | |

**Discussion**

Using protein-level data, we performed a large-scale proteomic analysis of 5,144 patient samples across 19 cancer types for β-Catenin and E-Cadherin. We found a high correlation between the β-Catenin and E-Cadherin across most cancers (**Figure 27A**), suggesting the two proteins likely function at the level of the adherens junction. β-Catenin and E-Cadherin were not highly correlated in in neural tumors, which do not express E-Cadherin in normal tissues, adrenocortical tumors, and renal papillary cell carcinoma. Utilizing this large data set, we were then able to ask the question of the prognostic value of β-Catenin and E-Cadherin (**Figure 27B**,**C**). Several observations countervailed previous IHC reports. For example, despite having a modest correlation between β-Catenin and E-Cadherin, KIRC tumors demonstrated good prognostic value for both factors (**Figure 27B**,**C**). This finding is refuted in the literature, with the absence of E-Cadherin in KIRC being reported as a poor prognostic indicator and the reduced-to-negative IHC staining of E-Cadherin can be used to differentiate KIRC from renal chromophobe tumors.[406,407]

Similar to KIRC, E-Cadherin in breast carcinoma has been used as both a differentiating marker and an indicator of poor prognosis.[373,384,393,395,396] Stratifying the TCGA BRCA cohort by histological diagnosis, we found E-Cadherin protein level predicted worse overall survival in IDC, but not in ILC. Likewise, in the TCGA BRCA cohort, an amalgamated patient cohort from Kaplan-Meier Plotter,[400] and the METABRIC breast cancer patient cohort,[401] we found the highest levels of RNA for *CDH1* was predictive of poor overall or recurrence-free survival (**Figure 29**). The poor prognostic indication of *CDH1* was also seen after stratifying IDC samples into $ER^+$ and $HER^+$ subgroups (**Figure 31C**). With only survival information available for 30 of the 163

patients from our IHC cohort, we were unable to address the prognostic implications. However, our IHC findings are similar to several studies finding a decrease in E-Cadherin predictive of tumor stage in IDC.[382,383]

This discordance between immunohistochemical and multi-omic analyses is not uncommon and has been the center of several concerted efforts to find greater agreement between the modalities in order to improve patient diagnostics and treatment.[408,409] Analogously, large-scale meta-analyses of the predictive value of E-Cadherin immunohistochemistry have been conducted in a number of cancers.[410–415] In general, these meta-analyses found loss of E-Cadherin as a poor prognostic indicator for overall and recurrence/progression-free survival, with the exception of colorectal cancer.[414] The coalescing of IHC-based studies into meta-analysis relies on correcting for differing methodologies of the immunohistology and is semi-quantitative. Conversely, RPPA quantifies over 200 proteins using validated antibodies via a micro-to-nano scale dot blot system.[416] Several limitations exist for RPPA in comparison to IHC in tumor pathology, centering on the loss of spatial distribution of the epitope and the sampling position effects, similar to RNA quantification.[417] In a similar avenue to sample positioning, impurity of the sample for tumor cells can have biased effects. Within the TCGA cohort, tumor cellular purity for the majority of cohorts range from 75-90%, with notable lower levels (>50%) in purity for lung adenocarcinoma (LUAD), stomach adenocarcinoma (STAD) and pancreatic adenocarcinoma (PAAD) cohorts.[418,419]

Immunohistochemical- and RNA-based subtyping of IDC are an established and growing component in patient management, respectively. Subtype-specific patterns in β-Catenin and E-Cadherin have previously found for E-Cadherin staining, with maintained E-Cadherin in Luminal/ER$^+$ tumors[384] and decreased E-Cadherin in TNBC.[420] This appears to be consistent with the normal mammary duct, with the luminal compartment staining positive for E-Cadherin, while the basal compartment has a variable staining pattern for E-Cadherin.[44] Beyond subtypic trends in E-Cadherin, we found a negative correlation between E-Cadherin and N-Cadherin in IDC samples (**Figure 28A**). Cadherin switching, from E-Cadherin to N-Cadherin, appears to not only mark post-EMT tumor cells, but leads to N-Cadherin-mediated cellular motility.[421,422] No coherent explanation was forthcoming for the poor survival at the highest levels of E-Cadherin mRNA or protein across IDC samples or across clinical subtypes.

In contrast, β-Catenin signaling has been reported as a poor prognostic marker independent of subtype and drives TNBC tumors.[405,423–425] Despite the reports of β-Catenin in promoting progression, positive nuclear localization via IHC of β-Catenin in IDC and ILC has been consistently uncommon.[405,426,427] This trend was recapitulated in our IHC cohort with only 2 of 163 patients positive for nuclear β-Catenin, one from the TNBC and the ER$^+$ subtypes. The strong protein correlation between β-Catenin and E-Cadherin, as well as the discoordination between β-Catenin protein and mRNA, drives us to the conclusion that structural β-Catenin

rather than WNT-activated nuclear β-Catenin drives the prognostic value in IDC (**Figure 28D**). We did notice subtype-specific difference of β-Catenin and E-Cadherin ratio (**Figure 31B**), indicating a more complex interaction between the two proteins at structural or signaling levels. For example, we found a relative reduction in β-Catenin compared to E-Cadherin in HER2$^+$ samples by IHC (**Table 3**) and in the RPPA data (**Figure 31B**). A recent article from *Tung et al.* found that monoallelic loss of *Ctnnb1*, the gene which encodes β-Catenin, drives an increase in HER2$^+$ murine tumorigenic capcity.[428] Monoallelic loss of *Ctnnb1* drove multiplicity in the HER2$^+$ mice and appeared to reduce differentiation by histological assessment, with the majority of tumors described as poorly-differentiated acini.[428]

In conclusion, our proteomic analysis revealed both β-Catenin and E-Cadherin are poor prognostic indicators in infiltrating ductal carcinoma. Similarly, across the TCGA BRCA dataset and multiple patient cohorts, *CDH1* gene expression was a poor prognosis indicator. This poor prognostic trend of *CDH1* was maintained across ER$^+$, HER2$^+$, and TNBC samples. The presence of β-Catenin and E-Cadherin in IDC patient samples was dependent on clinical subtype and stage, which may be a partial explanation for the difference in patient outcomes. Further assessment of E-Cadherin as a predictive marker in breast cancer is warranted, with pointed appraisal between emergent proteomic modalities and immunohistochemistry.

TRGAted: A WEB TOOL FOR SURVIVAL ANALYSIS USING PROTEIN DATA IN THE CANCER GENOME ATLAS

**Rationale**

Improving prognostic predictions and the identification of potential therapeutic targets is of particular interest to clinicians. Quantification of messenger RNA levels at a genome-wide level has proven valuable in the discovery of gene expression profiles, which can serve as biomarkers for clinical outcomes in cancer.[429] However, RNA quantification of tumor or patient cohorts is a proxy for protein level, with many cellular processes above transcription that ultimately regulate protein level. The availability of protein-level quantification for the TCGA cohorts allow for more relevant clinical outcome predictions compared to mRNA levels. Currently available applications provide entry-level analysis in correlational, differential, and survival modalities for the RPPA information. However, survival analysis in these applications rely on median- or mean-based survival data and do not allow for the use of clinical variables.[398,430,431]

With these limitations in mind, we developed a new open-source web application, TRGAted (**Figure 32**). Built on the R shiny framework, TRGAted is an intuitive data analysis tool for parsing survival information based on over 200 proteins in 31 cancer types. TRGAted is comprised of processed RPPA information, survival information, and code, allowing users to run instances locally or modify the code with ease.

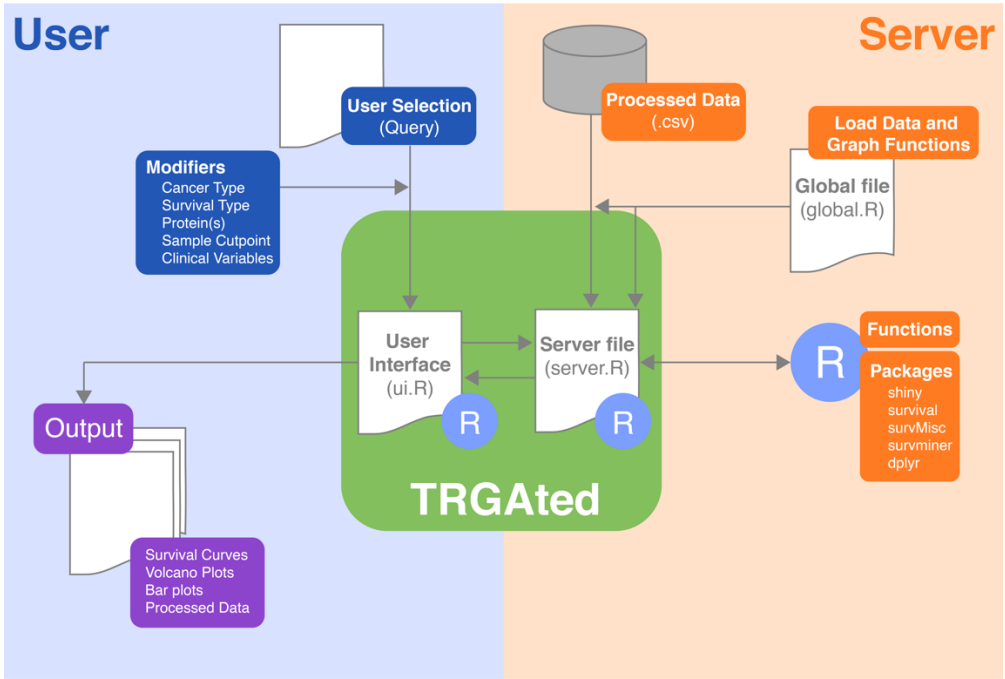


**Figure a.7.** Diagram of the implementation of TRGAted. Each file communicates within the R Shiny framework. On the user side (left, blue), users select pertinent cancer type, protein of interest, and clinical variables into the CSS-enabled user interface. This information is received by the server file enabling the subsequent run in R. On the server side (right, orange), the specific cancer type from the database, R packages, and functions are retrieved and executed. After execution, the server file provides both tabular and graphical output (purple) to the user interface and is displayed.

**Materials and Methods**

*Protein and survival data*

Level 4 TCGA RPPA data for each cancer type was downloaded from the TCPA Portal developed by the MD Anderson Cancer Center.[398] Across all proteins, individual values were scaled using Z-scores. A summary of information available for each cancer datasets is in **Table 5**. Additionally, uveal melanoma (UVM) was excluded from the datasets due to a low number of samples with RPPA quantification (n=12). Clinical and survival information for each cancer data set were downloaded from recently updated TCGA clinical data.[432] Of the 8,167 samples available in the TCPA, overall survival (OS) data was available for 7,714 patients, disease-specific survival (DSS) data was available for 7,240 patients, disease-free interval (DFI) data was available for 3,887 patients, and progression-free interval (PFI) data was available for 7,315 patients (**Table 5**). Unlike other cancer types, metastatic samples were kept in the skin cutaneous melanoma (SKCM) RPPA-based dataset due to the highly metastatic nature of the disease. SKCM in the TRGAted application consists of 96 primary tumor samples and 258 metastatic samples.

*Implementation*

The TRGAted application was written and tested using R (v3.5.1). The interactive plots are made using shiny (v1.1.0) and ggplot2 (v3.0.0). Plots can be downloaded as .png, .pdf, or .svg files. Data used to generate the individual plots can be downloaded as .csv files. Minimum system requirements for running TRGAted locally are modest and include an Intel-compatible CPU and 1 gigabyte of RAM. Running TRGAted from the shiny server requires a modern browser and an internet connection.

Kaplan-Meier survival curves can be generated by selecting the cancer type, survival type and protein(s) of interest (**Figure 33**). Kaplan-Meier curves are generated using the survival (v2.41-3) and the survminer (v0.4-1) R packages. Multi-protein survival analysis utilizes mean values of protein probes, similar to gene-expression-based survival analysis platforms.[404] Hazard ratio for two-group comparisons, either median or optimal cut-off, utilize the Cox proportional hazards regression model in the survival R package; with the reported hazard ratio comparing high *versus* low protein groups. Optimal cut-off feature uses the surv_cutpoint function of the survminer package, calculating the minimal p-value based on the log-rank method. This function uses the maximally selected rank statistic (maxstat, v0.7-25) R package, which finds the maximal standardized two-sample linear rank statistic.[433] In order to find clinically or biologically meaningful biomarkers, the minimal proportion cutpoint, or the maximal disparity comparison, was set at 15% *versus* 85% of samples. Clinical variables dependent on the cancer type selected, can be used to filter patients into user-defined groupings. Clinical information available across all types include: subtype, tumor stage, histological type, gender, age, and response to primary therapy.

**Table 5.** Summary of cancer types and survival information available in TRGAted.

| Cancer Type | Samples | OS | DSS | DFI | PFI | Proteins |
|---|---|---|---|---|---|---|
| Adrenocortical carcinoma (ACC) | 46 | 46 | 46 | 28 | 46 | 221 |
| Bladder Urothelial Carcinoma (BLCA) | 344 | 344 | 330 | 153 | 344 | 223 |
| Breast invasive carcinoma (BRCA) | 901 | 873 | 855 | 750 | 873 | 224 |
| Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) | 171 | 171 | 168 | 112 | 171 | 220 |
| Cholangiocarcinoma (CHOL) | 30 | 30 | 29 | 21 | 30 | 219 |
| Colon adenocarcinoma (COAD) | 358 | 325 | 311 | 126 | 325 | 223 |
| Diffuse Large B-cell Lymphoma (DLBCL) | 33 | 33 | 33 | 19 | 33 | 219 |
| Esophageal carcinoma (ESCA) | 126 | 126 | 124 | 76 | 126 | 220 |
| Glioblastoma multiforme (GBM) | 205 | 136 | 123 | 0 | 136 | 223 |
| Head and Neck squamous cell carcinoma (HNSC) | 346 | 346 | 326 | 85 | 346 | 239 |
| Kidney Chromophobe (KICH) | 63 | 63 | 63 | 27 | 63 | 220 |
| Kidney renal clear cell carcinoma (KIRC) | 445 | 444 | 434 | 72 | 444 | 233 |
| Kidney renal papillary cell carcinoma (KIRP) | 208 | 207 | 205 | 127 | 207 | 221 |
| Lower Grade Glioma (LGG) | 427 | 426 | 420 | 114 | 426 | 220 |
| Liver hepatocellular carcinoma (LIHC) | 184 | 184 | 177 | 145 | 184 | 220 |
| Lung adenocarcinoma (LUAD) | 362 | 361 | 327 | 203 | 361 | 239 |
| Lung squamous cell carcinoma (LUSC) | 325 | 325 | 295 | 210 | 325 | 239 |
| Mesothelioma (MESO) | 61 | 61 | 45 | 10 | 61 | 220 |
| Ovarian serous cystadenocarcinoma (OV) | 411 | 405 | 377 | 199 | 407 | 224 |
| Pancreatic adenocarcinoma (PAAD) | 105 | 105 | 99 | 40 | 105 | 221 |
| Pheochromocytoma and Paraganglioma (PCPG) | 81 | 79 | 79 | 71 | 79 | 220 |
| Prostate adenocarcinoma (PRAD) | 351 | 351 | 350 | 233 | 351 | 220 |
| Rectum adenocarcinoma (READ) | 130 | 126 | 120 | 31 | 126 | 223 |
| Sarcoma (SARC) | 221 | 221 | 215 | 125 | 22 | 220 |
| Skin Cutaneous Melanoma (SKCM) | 354 | 349 | 346 | 0 | 349 | 223 |
| Stomach adenocarcinoma (STAD) | 392 | 357 | 334 | 207 | 357 | 220 |
| Testicular Germ Cell Tumors (TGCT) | 118 | 104 | 104 | 79 | 104 | 219 |
| Thyroid carcinoma (THCA) | 374 | 372 | 366 | 268 | 372 | 219 |
| Thymoma (THYM) | 90 | 90 | 90 | 9 | 90 | 219 |
| Uterine Corpus Endometrial Carcinoma (UCEC) | 404 | 404 | 403 | 325 | 404 | 223 |
| Uterine Carcinosarcoma (UCS) | 48 | 48 | 46 | 22 | 48 | 220 |

TRGAted also allows for Cox proportional hazard modeling across all proteins in each cancer type or for a single protein across all cancer types. Hazard ratios and P values are based on the Cox regression model. Values filtered from the volcano plots are proteins with –log10(p-values) less than 0.1 and hazard ratios greater than 20. These filters were implemented to improve visualization and to reduce artifacts of the analysis pipeline, respectively. The volcano plot can be graphed as linear or natural-log transformed, to assist in the visualization of good prognostic indicators. Visualizing the proportional comparison for the volcano plots is also available.

*Use Case*

In order to demonstrate the functionality of TRGAted, we present a basic survival analysis of examining the aggressive, highly-metastatic subtype of breast cancer, known as basal-like breast cancer. We found in this cancer, RAD50, involved in homologous recombination of DNA, as a novel poor prognostic marker.

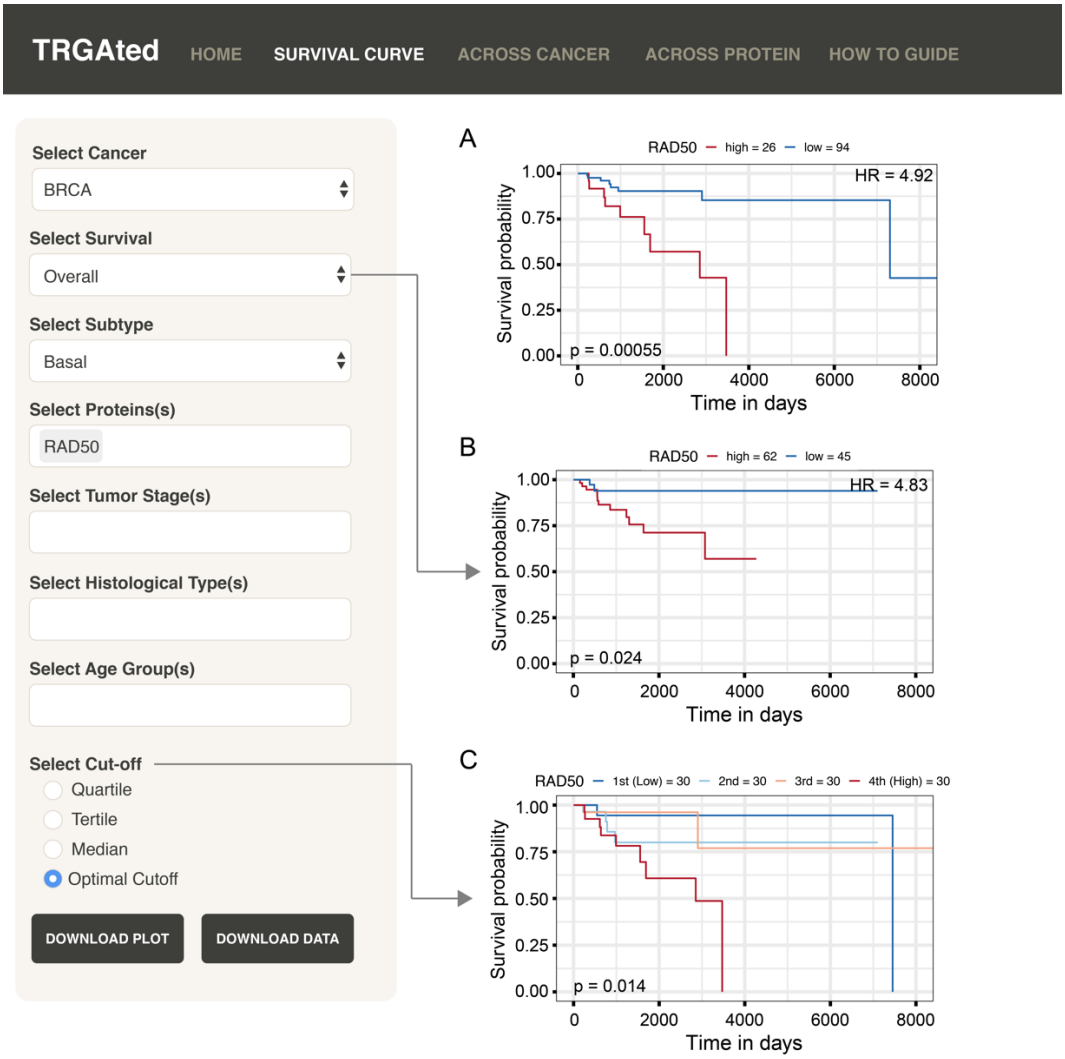


**Figure a.8.** Generating survival curves. The interface shows an example of an overall survival curve for the RAD50 protein in the basal subtype of breast cancer using the optimal cutpoint **A**. Disease-specific survival, disease-free interval, and progression-free interval can also be selected **B.** The cutpoint can be varied to separate samples based on protein level into quartiles, tertiles, medians or separating into two groups based on the lowest p-value **C.**

**Survival curves:** Survival curves can be generated by selecting the cancer type, survival type, and protein or proteins of interest (**Figure 33A**). We also selected the subtype information to more closely examine basal-like breast cancer. Other survival types and clinical variables can be selected (**Figure 33B**). The division of samples is available into quartiles, tertiles, median or optimum based on the protein of interest (**Figure 33C**). Here we can see that the DNA repair protein, RAD50 is a poor prognostic marker for overall (**Figure 33A**) and disease-specific survival (**Figure 33B**) in basal-like breast cancer.

**Across cancer:** TRGAted can be used for biomarker discovery by examining the hazard ratios for all proteins available by cancer subtype, like basal-like breast cancer (**Figure 34A**). The volcano plot displays good prognostic markers on the left in blue and poor prognostic markers on the right in red. Having selected the

optimal cutoff feature, a bar chart can also be generated to examine the proportion of samples in the high and low proportion groups (**Figure 34B**). Protein labeling is adaptive for both the volcano plot and bar chart and will only label significant proteins. Here we see the RAD50 is one of the most significant predictors of poor overall survival in basal-like breast cancer (**Figure 34A**,**B**).

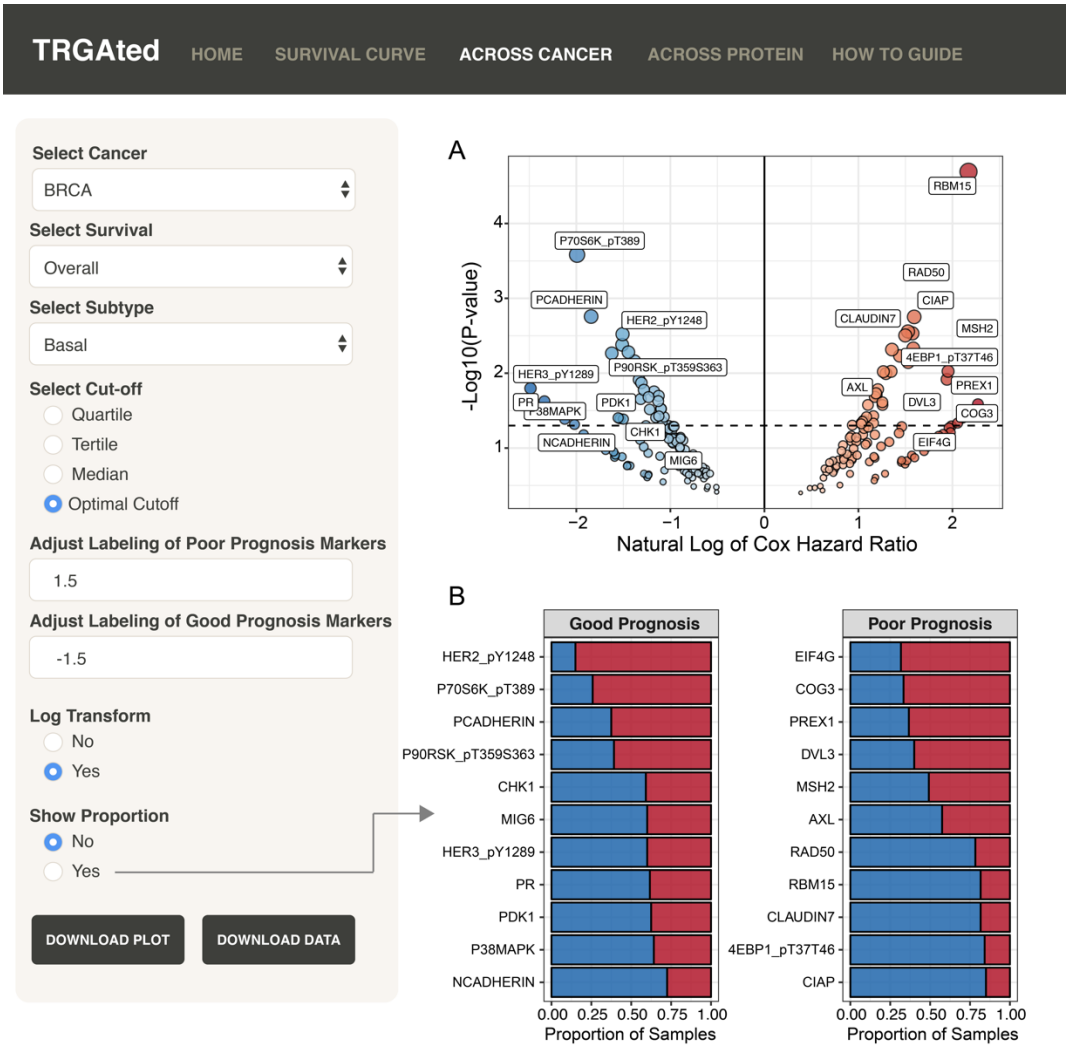


**Figure a.9.** Visualizing all proteins across a single cancer type. The interface shows an example of the visualization of Cox hazard ratio of each protein across the basal subtype of breast cancer **A.** Good prognostic markers appear on the left in blue, while poor prognostic markers are on the right in red. The natural log transformation allows the graph to be centered at 0 and makes the visualization of good prognostic markers easier. Labeling for proteins can be adjusted to include more or less protein. Proportional comparisons for protein using the optimal cutpoint function is available as well **B.**

**Across protein:** TRGAted can also be used to examine the survival outcomes of a protein of interest across multiple cancers. Here, RAD50 predicts poor survival in only five cancer types, prostate, adrenocortical, breast cancer, low-grade glioma and head and neck cancers (**Figure 35A**). A summary of the hazard ratios can also be visualized by selecting for the barplot function (**Figure 35B**).

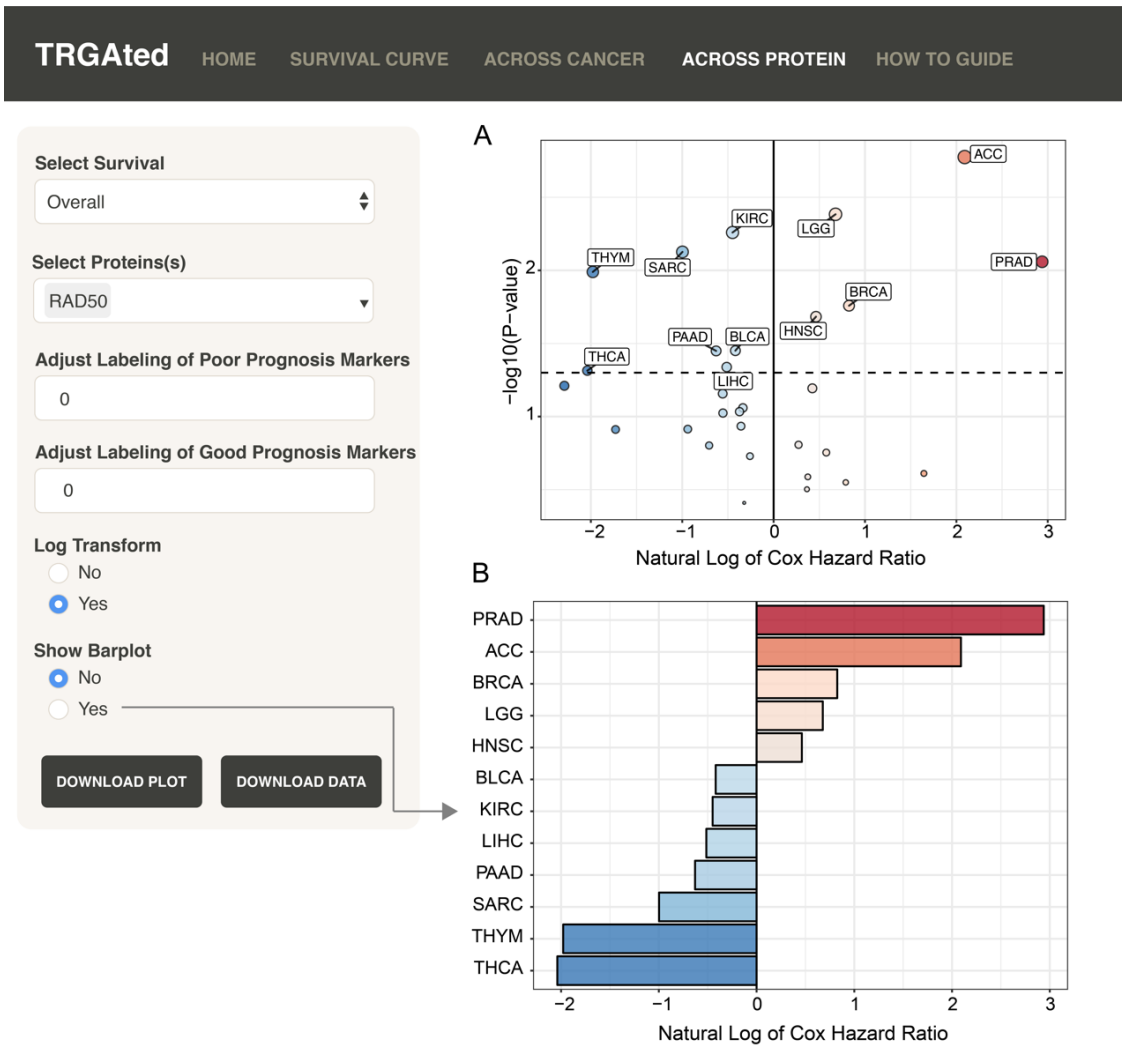


**Figure a.10.** Visualizing a protein across cancer types. **A.** The interface shows an example of the visualization of Cox hazard ratio of for RAD50 across all 31 cancer types. This feature is similar to the Across Cancer tab with the ability to adjust labels and log-transform the Cox hazard ratios. **B.** The hazard ratios for significant cancer types can be visualized using a bar chart (**B**).

## Conclusions

TRGAted is an open-source survival analysis application designed to allow for quick and intuitive exploration of TCGA protein-level data. This survival analysis improves on current TCGA pipelines by providing greater diversity of clinical and survival options and relying on protein-level data. In addition to log-rank and Cox regression modeling, TRGAted allows users to download graphical displays and processed data for up to 7,714 samples across 31 cancer types. Built on the R Shiny framework, a literate code architecture, the code for TRGAted is annotated and easily modified from our GitHub repository. Under the GNU General Public License v3.0, we encourage interested groups to modify TRGAted for greater usability. Downloading and modifying TRGAted is streamlined by the relatively small size of TRGAted, totally 27.2 megabytes for the application, processed data, and built-in instructional guide.

## REFERENCES


1. Doyle, A. C. *The sign of four*. (Broadview Press, 2010).
2. Taleb, N. N. *The Black Swan*. *Penguin Books Ltd* (2007). doi:10.5465/AMP.2011.61020810
3. Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. *Science (80-. ).* **211,** 453–458 (1981).
4. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA. Cancer J. Clin.* **66,** 7–30 (2016).
5. Nelson, H. D. *et al.* Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. Rockville, MD: Agency for Healthcare Research and Quality; 2009. (2014).
6. Kim, H. L., Puymon, M. R., Qin, M., Guru, K. & Mohler, J. L. NCCN clinical practice guidelines in oncology™. (2014).
7. Rosenberg, R. D. *et al.* Performance Benchmarks for Screening Mammography. *Radiology* **241,** 55–66 (2006).
8. Jørgensen, K. J. & Gøtzsche, P. C. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* **339,** b2587 (2009).
9. Cowell, C. F. *et al.* Progression from ductal carcinoma in situto invasive breast cancer: Revisited. *Mol. Oncol.* **7,** 859–869 (2013).
10. American Cancer Society & Society, A. C. Cancer Facts & Figures 2016. *Cancer Facts Fig. 2016* 1–9 (2016). doi:10.1097/01.NNR.0000289503.22414.79
11. Vu-Nishino, H., Tavassoli, F. A., Ahrens, W. A. & Haffty, B. G. Clinicopathologic features and long-term outcome of patients with medullary breast carcinoma managed with breast-conserving therapy (BCT). *Int. J. Radiat. Oncol. Biol. Phys.* **62,** 1040–1047 (2005).
12. Rakha, E. A. *et al.* Tubular carcinoma of the breast: Further evidence to support its excellent prognosis. *J. Clin. Oncol.* **28,** 99–104 (2010).
13. Pal, S. K. *et al.* Papillary carcinoma of the breast: An overview. *Breast Cancer Research and Treatment* **122,** 637–645 (2010).
14. Bae, S. Y. *et al.* Mucinous Carcinoma of the Breast in Comparison with Invasive Ductal Carcinoma: Clinicopathologic Characteristics and Prognosis. *J. Breast Cancer* **14,** 308 (2011).
15. Anderson, W. F., Schairer, C., Chen, B. E., Hance, K. W. & Levine, P. H. Epidemiology of inflammatory breast cancer (IBC). *Breast Dis.* **22,** 9–23 (2010).
16. Slamon, D. J. *et al.* Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (80-. ).* **235,** 177–182 (1987).
17. Slamon, D. J. *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science (80-. ).* **244,** 707–712 (1989).
18. Press, M. F. *et al.* Her-2/neu Expression in Node-negative Breast Cancer: Direct Tissue Quantitation by Computerized Image Analysis and Association of Overexpression with Increased Risk of Recurrent Disease. *Cancer Res.* **53,** 4960–4970 (1993).
19. Dent, R. *et al.* Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clin. Cancer Res.* **13,** 4429–4434 (2007).

20. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).

21. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98,** 10869–10874 (2001).

22. Badve, S. *et al.* Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology* **24,** 157–167 (2011).

23. Nielsen, T. O. *et al.* Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma Immunohistochemical and Clinical Characterization of the Basal- Like Subtype of Invasive Breast Carcinoma. *Clin. Cancer Res.* **10,** 5367–5374 (2004).

24. Bernard, P. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).

25. Tibshirani, R. *et al.* Diagnosis of Multiple Cancer Types by Shrunken Centroids of Gene Expression Diagnosis of multiple cancer centroids of gene expression. *Source Proc. Natl. Acad. Sci. United States Am.* **99,** 6567–6572 (2002).

26. Boccia, R. V. Translating research into practice: The prosigna® (PAM50) gene signature assay. *Clin. Adv. Hematol. Oncol.* **13,** 3–13 (2015).

27. Moore, G. E. Cramming more components onto integrated circuits. *Proc. IEEE* **86,** 82–85 (1998).

28. Hayden, E. C. The $ 1,000 genome. *Nature* **507,** 294–295 (2014).

29. O'Keefe, B. The future of cancer genomics. *Nat. Med.* **21,** 99–99 (2015).

30. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

31. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12,** (2010).

32. Lehmann, B. D. B. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121,** 2750–2767 (2011).

33. Miller, K. *et al.* Bevacizumab (Bv) in the adjuvant treatment of HER2-negative breast cancer: Final results from Eastern Cooperative Oncology Group E5103. *J. Clin. Oncol.* **32,** 500 (2014).

34. Cameron, D. *et al.* Adjuvant bevacizumab-containing therapy in triple-negative breast cancer (BEATRICE): Primary results of a randomised, phase 3 trial. *Lancet Oncol.* **14,** 933–942 (2013).

35. Carey, L. A. *et al.* TBCRC 001: EGFR inhibition with cetuximab added to carboplatin in metastatic triple-negative (basal-like) breast cancer. *J. Clin. Oncol.* **26,** 1009–1009 (2008).

36. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N. Engl. J. Med.* **377,** 523–533 (2017).

37. Rouleau, M., Patel, A., Hendzel, M. J., Kaufmann, S. H. & Poirier, G. G. PARP inhibition: PARP1 and beyond. *Nature Reviews Cancer* **10,** 293–301 (2010).

38. Antoniou, A. *et al.* Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies. *Am. J. Hum. Genet.* **72,** 1117–1130 (2003).

39. O'Shaughnessy, J. *et al.* Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *N Engl J Med* **364,** 205–214 (2011).

40. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20,** 68–80 (2010).

41. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472,** 90–95 (2011).

42. Navin, N. E. The first five years of single-cell cancer genomics and beyond. *Genome Research* **25,** 1499–1507 (2015).

43. Chung, W. *et al.* Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* **8,** (2017).

44. Coley, W. B. Contribution to the knowledge of sarcoma. *Ann. Surg.* **14,** 199–200 (1891).

45. McCarthy, E. F. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop. J.* **26,** 154–8 (2006).

46. Couzin-Frankel, J. Cancer Immunotherapy. *Science (80-. ).* **342,** 1432–1433 (2013).

47. Kaiser, J. & Couzin-Frankel, J. Cancer immunotherapy sweeps Nobel for medicine. (2018).

48. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

49. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

50. Mellman, I. & Nussenzweig, M. Retrospective. Ralph M. Steinman (1943-2011). *Science* **334,** 466 (2011).

51. Smith-Garvin, J. E., Koretzky, G. A. & Jordan, M. S. T cell activation. *Annu. Rev. Immunol.* **27,** 591–619 (2009).

52. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck. *Cell* **55,** 301–308 (1988).

53. Barber, E. K., Dasgupta, J. D., Schlossman, S. F., Trevillyan, J. M. & Rudd, C. E. The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56lck) that phosphorylates the CD3 complex. *Proc Natl Acad Sci U S A* **86,** 3277–3281 (1989).

54. Bubeck Wardenburg, J. *et al.* Phosphorylation of SLP-76 by the ZAP-70 protein-tyrosine kinase is required for T-cell receptor function. *J Biol Chem* **271,** 19641–19644 (1996).

55. Zhang, W., Sloan-Lancaster, J., Kitchen, J., Trible, R. P. & Samelson, L. E. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell* **92,** 83–92 (1998).

56. Reynolds, L. F. *et al.* Vav1 transduces T cell receptor signals to the activation of phospholipase C-gamma1 via phosphoinositide 3-kinase-dependent and -independent pathways. *J Exp Med* **195,** 1103–1114 (2002).

57. Pages, F. *et al.* Binding of phosphatidylinositol-3-OH kinase to CD28 is required for T-cell signalling. *Nature* **369,** 327–329 (1994).

58. Acuto, O. & Michel, F. CD28-mediated co-stimulation: a quantitative support for TCR signalling. *Nat Rev Immunol* **3,** 939–951 (2003).

59. Zhu, J. & Paul, W. E. CD4 T cells: fates, functions, and faults. *Blood* **112,** 1557–1569 (2008).

60. Chen, L. & Flies, D. B. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol* **13,** 227–242 (2013).

61. Garrido, F. *et al.* Implications for immunosurveillance of altered HLA class I phenotypes in human tumours. *Immunol. Today* **18,** 89–95 (1997).

62. Garrido, F. & Algarra, I. MHC antigens and tumor escape from immune surveillance. (2001).

63. Drake, C. G., Jaffee, E. & Pardoll, D. M. Mechanisms of immune evasion by tumors. *Adv Immunol* **90,** 51–81 (2006).

64. de Visser, K. E., Eichten, A. & Coussens, L. M. Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer* **6,** 24–37 (2006).

65. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Curr Opin Immunol* **24,** 207–212 (2012).

66. Linsley, P. S. *et al.* Intracellular trafficking of CTLA-4 and focal localization towards sites of TCR engagement. *Immunity* **4,** 535–543 (1996).

67. Rudd, C. E., Taylor, A. & Schneider, H. CD28 and CTLA-4 coreceptor expression and signal transduction. *Immunol Rev* **229,** 12–26 (2009).

68. Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J Exp Med* **192,** 303–310 (2000).

69. Wing, K. *et al.* CTLA-4 control over Foxp3+ regulatory T cell function. *Science (80-. ).* **322,** 271–275 (2008).

70. Selby, M. J. *et al.* Anti-CTLA-4 Antibodies of IgG2a Isotype Enhance Antitumor Activity through Reduction of Intratumoral Regulatory T Cells. *Cancer Immunol. Res.* **1,** 32–42 (2013).

71. Vargas, F. A. *et al.* Fc effector function contributes to the activity of human anti-CTLA-4 antibodies. *Cancer Cell* **33,** 649–663. e4 (2018).

72. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* **12,** 252–264 (2012).

73. Khattri, R., Auger, J. A., Griffin, M. D., Sharpe, A. H. & Bluestone, J. A. Lymphoproliferative disorder in CTLA-4 knockout mice is characterized by CD28-regulated activation of Th2 responses. *J. Immunol.* **162,** 5784–5791 (1999).

74. Tivol, E. A. *et al.* CTLA4Ig prevents lymphoproliferation and fatal multiorgan tissue destruction in CTLA-4-deficient mice. *J. Immunol.* **158,** 5091–5094 (1997).

75. Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423,** 506–511 (2003).

76. Kristiansen, O. P., Larsen, Z. M. & Pociot, F. CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? *Genes Immun* **1,** 170–184 (2000).

77. Keir, M. E., Butte, M. J., Freeman, G. J. & Sharpe, A. H. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol* **26,** 677–704 (2008).

78. Blank, C. *et al.* Absence of programmed death receptor 1 alters thymic development and enhances generation of CD4/CD8 double-negative TCR-transgenic T cells. *J Immunol* **171,** 4574–4581 (2003).

79. Dong, H. *et al.* Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat Med* **8,** 793–800 (2002).

80. Wintterle, S. *et al.* Expression of the B7-related molecule B7-H1 by glioma cells: a potential mechanism of immune paralysis. *Cancer Res* **63,** 7462–7467 (2003).

81. Garcia-Diaz, A. *et al.* Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. *Cell Rep* **19,** 1189–1201 (2017).

82. Mimura, K. *et al.* PD-L1 expression is mainly regulated by interferon gamma associated with JAK-STAT pathway in gastric cancer. *Cancer Sci* **109,** 43–53 (2018).

83. Yamazaki, T. *et al.* Expression of programmed death 1 ligands by murine T cells and APC. *J Immunol* **169,** 5538–5545 (2002).

84. Schreiner, B. *et al.* Interferon-beta enhances monocyte and dendritic cell expression of B7-H1 (PD-L1), a strong inhibitor of autologous T-cell activation: relevance for the immune modulatory effect in multiple sclerosis. *J Neuroimmunol* **155,** 172–182 (2004).

85. Chemnitz, J. M., Parry, R. V, Nichols, K. E., June, C. H. & Riley, J. L. SHP-1 and SHP-2 associate with immunoreceptor tyrosine-based switch motif of programmed death 1 upon primary human T cell stimulation, but only receptor ligation prevents T cell activation. *J Immunol* **173,** 945–954 (2004).

86. Sheppard, K. A. *et al.* PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3zeta signalosome and downstream signaling to PKCtheta. *FEBS Lett* **574,** 37–41 (2004).

87. Chen, W. *et al.* Conversion of Peripheral CD4 + CD25 − Naive T Cells to CD4 + CD25 + Regulatory T Cells by TGF-β Induction of Transcription Factor Foxp3. *J. Exp. Med.* **198,** 1875–1886 (2003).

88. Francisco, L. M. *et al.* PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med* **206,** 3015–3029 (2009).

89. Nishimura, H., Nose, M., Hiai, H., Minato, N. & Honjo, T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* **11,** 141–151 (1999).

90. Nishimura, H. *et al.* Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. *Science (80-. ).* **291,** 319–322 (2001).

91. Wang, J. *et al.* PD-1 deficiency results in the development of fatal myocarditis in MRL mice. *Int Immunol* **22,** 443–452 (2010).

92. Ahmed, R. & Honjo, T. *Negative co-receptors and ligands.* **350,** (Springer Science & Business Media, 2011).

93. Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439,** 682–687 (2006).

94. Catakovic, K., Klieser, E., Neureiter, D. & Geisberger, R. T cell exhaustion: from pathophysiological basics to tumor immunotherapy. *Cell Commun Signal* **15,** 1 (2017).

95. Wherry, E. J. T cell exhaustion. *Nat Immunol* **12,** 492–499 (2011).

96. Kahan, S. M., Wherry, E. J. & Zajac, A. J. T cell exhaustion during persistent viral infections. *Virology* **479–480,** 180–193 (2015).

97. Jiang, Y., Li, Y. & Zhu, B. T-cell exhaustion in the tumor microenvironment. *Cell Death Dis* **6,** e1792 (2015).

98. Lu, C., Redd, P. S., Lee, J. R., Savage, N. & Liu, K. The expression profiles and regulation of PD-L1 in tumor-induced myeloid-derived suppressor cells. *Oncoimmunology* **5,** e1247135 (2016).

99. Nishida, N. & Kudo, M. Oncogenic Signal and Tumor Microenvironment in Hepatocellular Carcinoma. *Oncology* **93 Suppl 1,** 160–164 (2017).

100. Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science (80-. ).* **348,** 56–61 (2015).

101. Le Mercier, I., Lines, J. L. & Noelle, R. J. Beyond CTLA-4 and PD-1, the Generation Z of Negative Checkpoint Regulators. *Front Immunol* **6,** 418 (2015).

102. Fourcade, J. *et al.* Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients. *J Exp Med* **207,** 2175–2186 (2010).

103. Kuchroo, V. K., Umetsu, D. T., DeKruyff, R. H. & Freeman, G. J. The TIM gene family: emerging roles in immunity and disease. *Nat Rev Immunol* **3,** 454–462 (2003).

104. Goldberg, M. V & Drake, C. G. LAG-3 in Cancer Immunotherapy. *Curr Top Microbiol Immunol* **344,** 269–278 (2011).

105. Triebel, F. *et al.* LAG-3, a novel lymphocyte activation gene closely related to CD4. *J. Exp. Med.* **171,** 1393–1405 (1990).

106. Sehrawat, S. *et al.* Galectin-9/TIM-3 interaction regulates virus-specific primary and memory CD8 T cell response. *PLoS Pathog* **6,** e1000882 (2010).

107. Le Mercier, I. *et al.* VISTA Regulates the Development of Protective Antitumor Immunity. *Cancer Res* **74,** 1933–1944 (2014).

108. Li, M. *et al.* T-cell immunoglobulin and ITIM domain (TIGIT) receptor/poliovirus receptor (PVR) ligand engagement suppresses interferon-gamma production of natural killer cells via beta-arrestin 2-mediated negative signaling. *J Biol Chem* **289,** 17647–17657 (2014).

109. Liu, S. *et al.* Recruitment of Grb2 and SHIP1 by the ITT-like motif of TIGIT suppresses granule polarization and cytotoxicity of NK cells. *Cell Death Differ* **20,** 456–464 (2013).

110. Tahara-Hanaoka, S. *et al.* Tumor rejection by the poliovirus receptor family ligands of the DNAM-1 (CD226) receptor. *Blood* **107,** 1491–1496 (2006).

111. Stanietsky, N. *et al.* Mouse TIGIT inhibits NK-cell cytotoxicity upon interaction with PVR. *Eur J Immunol* **43,** 2138–2150 (2013).

112. Manieri, N. A., Chiang, E. Y. & Grogan, J. L. TIGIT: A Key Inhibitor of the Cancer Immunity Cycle. *Trends Immunol* **38,** 20–28 (2017).

113. Dong, C. *et al.* ICOS co-stimulatory receptor is essential for T-cell activation and function. *Nature* **409,** 97–101 (2001).

114. Linch, S. N., McNamara, M. J. & Redmond, W. L. OX40 Agonists and Combination Immunotherapy: Putting the Pedal to the Metal. *Front Oncol* **5,** 34 (2015).

115. Rogers, P. R., Song, J., Gramaglia, I., Killeen, N. & Croft, M. OX40 promotes Bcl-xL and Bcl-2 expression and is essential for long-term survival of CD4 T cells. *Immunity* **15,** 445–455 (2001).

116. Piconese, S., Valzasina, B. & Colombo, M. P. OX40 triggering blocks suppression by regulatory T cells and facilitates tumor rejection. *J Exp Med* **205,** 825–839 (2008).

117. Quezada, S. A., Jarvinen, L. Z., Lind, E. F. & Noelle, R. J. CD40/CD154 interactions at the interface of tolerance and immunity. *Annu Rev Immunol* **22,** 307–328 (2004).

118. Robert, C. *et al.* Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med* **372,** 2521–2532 (2015).

119. Yamazaki, N. *et al.* Phase 1b study of pembrolizumab (MK-3475; anti-PD-1 monoclonal antibody) in Japanese patients with advanced melanoma (KEYNOTE-041). *Cancer Chemother Pharmacol* **79,** 651–660 (2017).

120. Ribas, A. *et al.* Association of Pembrolizumab With Tumor Response and Survival Among Patients With Advanced Melanoma. *JAMA* **315,** 1600–1609 (2016).

121. Larkin, J. *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* **373,** 23–34 (2015).

122. Robert, C. *et al.* Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* **372,** 320–330 (2015).

123. Topalian, S. L. *et al.* Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. *N. Engl. J. Med.* **366,** 2443–2454 (2012).

124. Ascierto, P. A. *et al.* Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol* **18,** 611–622 (2017).

125. Hodi, F. S. *et al.* Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* **17,** 1558–1568 (2016).

126. Hamid, O. *et al.* A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma. *J Transl Med* **9,** 204 (2011).

127. Zimmer, L. *et al.* Open-label, multicenter, single-arm phase II DeCOG-study of ipilimumab in pretreated patients with different subtypes of metastatic melanoma. *J Transl Med* **13,** 351 (2015).

128. Herbst, R. S. *et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* **387,** 1540–1550 (2016).

129. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **375,** 1823–1833 (2016).

130. Garon, E. B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* **372,** 2018–2028 (2015).

131. Kato, T. *et al.* KEYNOTE-025: Phase 1b study of pembrolizumab (pembro) in Japanese patients (pts) with previously treated PD-L1+ non-small cell lung cancer (NSCLC). *Ann. Oncol.* **27,** (2016).

132. Rizvi, N. A. *et al.* Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol* **16,** 257–265 (2015).

133. Carbone, D. P. *et al.* First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N Engl J Med* **376,** 2415–2426 (2017).

134. Borghaei, H. *et al.* Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* **373,** 1627–1639 (2015).

135. Rittmeyer, A. *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* **389,** 255–265 (2017).

136. Fehrenbacher, L. *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387,** 1837–1846 (2016).

137. Petrylak, D. P. *et al.* Atezolizumab (MPDL3280A) Monotherapy for Patients With Metastatic Urothelial Cancer: Long-term Outcomes From a Phase 1 Study. *JAMA Oncol* (2018). doi:10.1001/jamaoncol.2017.5440

138. Powles, T. *et al.* Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* (2017). doi:10.1016/S0140-6736(17)33297-X

139. Balar, A. V *et al.* Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* **389,** 67–76 (2017).

140. Bellmunt, J. *et al.* Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N Engl J Med* **376,** 1015–1026 (2017).

141. Sharma, P. *et al.* Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol* **18,** 312–322 (2017).

142. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373,** 1803–1813 (2015).

143. Motzer, R. J. *et al.* Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. *J Clin Oncol* **33,** 1430–1437 (2015).

144. Armand, P. *et al.* Programmed Death-1 Blockade With Pembrolizumab in Patients With Classical Hodgkin Lymphoma After Brentuximab Vedotin Failure. *J Clin Oncol* **34,** 3733–3739 (2016).

145. Chen, R. *et al.* Phase II Study of the Efficacy and Safety of Pembrolizumab for Relapsed/Refractory Classic Hodgkin Lymphoma. *J Clin Oncol* **35,** 2125–2132 (2017).

146. Ansell, S. M. *et al.* PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med* **372,** 311–319 (2015).

147. Younes, A. *et al.* Nivolumab for classical Hodgkin's lymphoma after failure of both autologous stem-cell transplantation and brentuximab vedotin: a multicentre, multicohort, single-arm phase 2 trial. *Lancet Oncol* **17,** 1283–1294 (2016).

148. Kang, Y. K. *et al.* Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **390,** 2461–2471 (2017).

149. Ferris, R. L. *et al.* Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* **375,** 1856–1867 (2016).

150. El-Khoueiry, A. B. *et al.* Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* **389,** 2492–2502 (2017).

151. Tawbi, H. A. *et al.* Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol* **18,** 1493–1501 (2017).

152. D'Angelo, S. P. *et al.* Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. *Lancet Oncol.* (2018).

153. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372,** 2509–2520 (2015).

154. Giaccone, G. *et al.* Pembrolizumab in patients with thymic carcinoma: a single-arm, single-centre, phase 2 study. *Lancet Oncol* (2018). doi:10.1016/S1470-2045(18)30062-7

155. Morris, V. K. *et al.* Nivolumab for previously treated unresectable metastatic anal cancer (NCI9673): a multicentre, single-arm, phase 2 study. *Lancet Oncol* **18,** 446–453 (2017).

156. Nanda, R. *et al.* Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. *J Clin Oncol* **34,** 2460–2467 (2016).

157. Nghiem, P. T. *et al.* PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma. *N Engl J Med* **374,** 2542–2552 (2016).

158. Ding, W. *et al.* Pembrolizumab in patients with CLL and Richter transformation or with relapsed CLL. *Blood* **129,** 3419–3427 (2017).

159. Davids, M. S. *et al.* Ipilimumab for Patients with Relapse after Allogeneic Transplantation. *N Engl J Med* **375,** 143–153 (2016).

160. Zinzani, P. L. *et al.* Safety and tolerability of pembrolizumab in patients with relapsed/refractory primary mediastinal large B-cell lymphoma. *Blood* **130,** 267–270 (2017).

161. Yamazaki, N. *et al.* Phase 1b study of pembrolizumab (MK-3475; anti-PD-1 monoclonal antibody) in Japanese patients with advanced melanoma (KEYNOTE-041). *Cancer Chemother Pharmacol* **79,** 651–660 (2017).

162. Robert, C. *et al.* Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* **372,** 320–330 (2015).

163. Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* **366,** 2443–2454 (2012).

164. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* **375,** 1823–1833 (2016).

165. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-. ).* **348,** 124–128 (2015).

166. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45,** 228–247 (2009).

167. Armand, P. *et al.* Programmed Death-1 Blockade With Pembrolizumab in Patients With Classical Hodgkin Lymphoma After Brentuximab Vedotin Failure. *J Clin Oncol* **34,** 3733–3739 (2016).

168. Younes, A. *et al.* Nivolumab for classical Hodgkin's lymphoma after failure of both autologous stem-cell transplantation and brentuximab vedotin: a multicentre, multicohort, single-arm phase 2 trial. *Lancet Oncol* **17,** 1283–1294 (2016).

169. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (80-. ).* **357,** 409–413 (2017).

170. Nghiem, P. T. *et al.* PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma. *N Engl J Med* **374,** 2542–2552 (2016).

171. Royal, R. E. *et al.* Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *J Immunother* **33,** 828–833 (2010).

172. Brahmer, J. R. *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* **366,** 2455–2465 (2012).

173. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140,** 883–899 (2010).

174. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420,** 860–867 (2002).

175. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (80-. ).* **351,** 1463–1469 (2016).

176. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science (80-. ).* **350,** 207–211 (2015).

177. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **371,** 2189–2199 (2014).

178. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* **373,** 1803–1813 (2015).

179. Motzer, R. J. *et al.* Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. *J Clin Oncol* **33,** 1430–1437 (2015).

180. Miao, D. *et al.* Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science (80-. ).* **359,** 801–806 (2018).

181. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165,** 35–44 (2016).

182. Nolan, E. *et al.* Combined immune checkpoint blockade as a therapeutic strategy for BRCA1-mutated breast cancer. *Sci Transl Med* **9,** (2017).

183. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-. ).* **355,** (2017).

184. Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* **9,** (2017).

185. Luksza, M. *et al.* A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551,** 517–520 (2017).

186. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547,** 217–221 (2017).

187. Stronen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science (80-. ).* **352,** 1337–1341 (2016).

188. Linnemann, C. *et al.* High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med* **21,** 81–85 (2015).

189. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547,** 222–226 (2017).

190. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171,** 934–949 e15 (2017).

191. Riaz, N. *et al.* Recurrent SERPINB3 and SERPINB4 mutations in patients who respond to anti-CTLA4 immunotherapy. *Nat Genet* **48,** 1327–1329 (2016).

192. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* **375,** 819–829 (2016).

193. Lee, C. K. *et al.* Checkpoint inhibitors in metastatic EGFR-mutated non–small cell lung cancer—a meta-analysis. *J. Thorac. Oncol.* **12,** 403–407 (2017).

194. Taube, J. M. *et al.* Colocalization of inflammatory response with B7-h1 expression in human melanocytic lesions supports an adaptive resistance mechanism of immune escape. *Sci Transl Med* **4,** 127ra37 (2012).

195. Taube, J. M. *et al.* Differential Expression of Immune-Regulatory Genes Associated with PD-L1 Display in Melanoma: Implications for PD-1 Pathway Blockade. *Clin Cancer Res* **21,** 3969–3976 (2015).

196. Passiglia, F. *et al.* PD-L1 expression as predictive biomarker in patients with NSCLC: a pooled analysis. *Oncotarget* **7,** 19738–19747 (2016).

197. Weber, J. S. *et al.* Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* **16,** 375–384 (2015).

198. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515,** 568–571 (2014).

199. Inoue, H. *et al.* Intratumoral expression levels of PD-L1, GZMA, and HLA-A along with oligoclonal T cell expansion associate with response to nivolumab in metastatic melanoma. *Oncoimmunology* **5,** e1204507 (2016).

200. Taube, J. M. *et al.* Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. *Clin Cancer Res* **20,** 5064–5074 (2014).

201. Brahmer, J. *et al.* Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* **373,** 123–135 (2015).

202. Gandini, S., Massi, D. & Mandala, M. PD-L1 expression in cancer patients receiving anti PD-1/PD-L1 antibodies: A systematic review and meta-analysis. *Crit Rev Oncol Hematol* **100,** 88–98 (2016).

203. Lizotte, P. H. *et al.* Multiparametric profiling of non-small-cell lung cancers reveals distinct immunophenotypes. *JCI Insight* **1,** e89014 (2016).

204. Ribas, A. *et al.* Oncolytic Virotherapy Promotes Intratumoral T Cell Infiltration and Improves Anti-PD-1 Immunotherapy. *Cell* **170,** 1109–1119 e10 (2017).

205. Twyman-Saint Victor, C. *et al.* Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. *Nature* **520,** 373–377 (2015).

206. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science (80-. ).* **350,** 207–211 (2015).

207. Gao, J. *et al.* Loss of IFN-gamma Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell* **167,** 397–404 e9 (2016).

208. Daud, A. I. *et al.* Tumor immune profiling predicts response to anti-PD-1 therapy in human melanoma. *J Clin Invest* **126,** 3447–3452 (2016).

209. Smyrk, T. C., Watson, P., Kaul, K. & Lynch, H. T. Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma. *Cancer* **91,** 2417–2422 (2001).

210. Lynch, H. T., Drescher, K. M. & de la Chapelle, A. Immunology and the Lynch syndrome. *Gastroenterology* **134,** 1246–1249 (2008).

211. Llosa, N. J. *et al.* The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov.* **5,** 43–51 (2015).

212. Soares, K. C. *et al.* PD-1/PD-L1 blockade together with vaccine therapy facilitates effector T-cell infiltration into pancreatic tumors. *J Immunother* **38,** 1–11 (2015).

213. Schmittnaegel, M. *et al.* Dual angiopoietin-2 and VEGFA inhibition elicits antitumor immunity that is enhanced by PD-1 checkpoint blockade. *Sci Transl Med* **9,** (2017).

214. Sharabi, A. B., Lim, M., DeWeese, T. L. & Drake, C. G. Radiation and checkpoint blockade immunotherapy: radiosensitisation and potential mechanisms of synergy. *Lancet Oncol* **16,** e498-509 (2015).

215. Takahashi, Y. *et al.* Radiation enhanced the local and distant anti-tumor efficacy in dual immune checkpoint blockade therapy in osteosarcoma. *PLoS One* **12,** e0189697 (2017).

216. Lan, Y. *et al.* Enhanced preclinical antitumor activity of M7824, a bifunctional fusion protein simultaneously targeting PD-L1 and TGF-beta. *Sci Transl Med* **10,** (2018).

217. Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162,** 974–986 (2015).

218. Spranger, S. *et al.* Up-regulation of PD-L1, IDO, and Tregs in the melanoma tumor microenvironment is driven by CD8+ T cells. *Sci. Transl. Med.* **5,** 200ra116-200ra116 (2013).

219. Holmgaard, R. B., Zamarin, D., Munn, D. H., Wolchok, J. D. & Allison, J. P. Indoleamine 2,3-dioxygenase is a critical resistance mechanism in antitumor T cell immunotherapy targeting CTLA-4. *J Exp Med* **210,** 1389–1402 (2013).

220. Pitt, J. M. *et al.* Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrinsic Factors. *Immunity* **44,** 1255–1269 (2016).

221. Loke, P. & Allison, J. P. PD-L1 and PD-L2 are differentially regulated by Th1 and Th2 cells. *Proc Natl Acad Sci U S A* **100,** 5336–5341 (2003).

222. Brahmer, J. R. *et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J Clin Oncol* **28,** 3167–3175 (2010).

223. Robert, L. *et al.* Distinct immunological mechanisms of CTLA-4 and PD-1 blockade revealed by analyzing TCR usage in blood lymphocytes. *Oncoimmunology* **3,** e29244 (2014).

224. Simeone, E. *et al.* Immunological and biological changes during ipilimumab treatment and their potential correlation with clinical response and survival in patients with advanced melanoma. *Cancer Immunol Immunother* **63,** 675–683 (2014).

225. Delyon, J. *et al.* Experience in daily practice with ipilimumab for the treatment of patients with metastatic melanoma: an early increase in lymphocyte and eosinophil counts is associated with improved survival. *Ann. Oncol.* **24,** 1697–1703 (2013).

226. Krieg, C. *et al.* High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat. Med.* (2018).

227. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science (80-. ).* **359,** 582–587 (2018).

228. Wolchok, J. D. *et al.* Ipilimumab efficacy and safety in patients with advanced melanoma: a retrospective analysis of HLA subtype from four trials. *Cancer Immun* **10,** 9 (2010).

229. Johnson, D. B. *et al.* Fulminant Myocarditis with Combination Immune Checkpoint Blockade. *N Engl J Med* **375,** 1749–1755 (2016).

230. Postow, M. A. *et al.* Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N Engl J Med* **372,** 2006–2017 (2015).

231. Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science (80-. ).* **359,** 91–97 (2018).

232. Vetizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science (80-. ).* **350,** 1079–1084 (2015).

233. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science (80-. ).* **359,** 97–103 (2018).

234. Matson, V. *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science (80-. ).* **359,** 104–108 (2018).

235. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science (80-. ).* **350,** 1084–1089 (2015).

236. Berman, D. *et al.* Blockade of cytotoxic T-lymphocyte antigen-4 by ipilimumab results in dysregulation of gastrointestinal immunity in patients with advanced melanoma. *Cancer Immun* **10,** 11 (2010).

237. Dubin, K. *et al.* Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. *Nat Commun* **7,** 10391 (2016).

238. Kostic, A. D., Xavier, R. J. & Gevers, D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* **146,** 1489–1499 (2014).

239. Kroemer, G. & Zitvogel, L. Cancer immunotherapy in 2017: The breakthrough of the microbiota. *Nat Rev Immunol* **18,** 87–88 (2018).

240. Network, T. C. G. A. Comprehensive molecular portraits of human breast tumors. *Nature* **490,** 61–70 (2012).

241. Sørlie, T. & Perou, C. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. …* **98,** 10869–74 (2001).

242. Cheang, M. C. U. *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin. Cancer Res.* **14,** 1368–1376 (2008).

243. Marmé, F. & Schneeweiss, A. Targeted Therapies in Triple-Negative Breast Cancer. *Breast Care* **10,** 159–166 (2015).

244. Weigman, V. J. *et al.* Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res. Treat.* **133,** 865–880 (2012).

245. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. *Nature Reviews Clinical Oncology* **13,** 674–690 (2016).

246. Foulkes, W. D. Germline BRCA1 Mutations and a Basal Epithelial Phenotype in Breast Cancer. *CancerSpectrum Knowl. Environ.* **95,** 1482–1485 (2003).

247. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–399 (2012).

248. Miki, Y. *et al.* Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. *Science (80-. ).* **266,** 66–71 (1994).

249. KB, K., JL, H., DR, B. & Al, E. Risks of breast, ovarian, and contralateral breast cancer for brca1 and brca2 mutation carriers. *JAMA* **317,** 2402–2416 (2017).

250. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline *BRCA* Mutation. *N. Engl. J. Med.* **377,** 523–533 (2017).

251. Ledermann, J. *et al.* Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *N. Engl. J. Med.* **366,** 1382–1392 (2012).

252. Pahuja, S. *et al.* Outcome of BRCA 1/2-mutated (BRCA+) and triple-negative, BRCA wild type (BRCA-wt) breast cancer patients in a phase I study of single-agent veliparib (V). *J. Clin. Oncol.* **32,** 135 (2014).

253. Gelmon, K. A. *et al.* Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: A phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol.* **12,** 852–861 (2011).

254. Yamamoto, N. *et al.* A phase I, dose-finding and pharmacokinetic study of olaparib (AZD2281) in Japanese patients with advanced solid tumors. *Cancer Sci.* **103,** 504–509 (2012).

255. Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75,** 1215–1225 (1993).

256. Kunkel, T. A. & Erie, D. A. DNA MISMATCH REPAIR. *Annu. Rev. Biochem.* **74,** 681–710 (2005).

257. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372,** 2509–2520 (2015).

258. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (80-. ).* **357,** 409–413 (2017).

259. Borcherding, N. *et al.* Keeping Tumors in Check: A Mechanistic Review of Clinical Response and Resistance to Immune Checkpoint Blockade in Cancer. *Journal of Molecular Biology* (2018). doi:10.1016/j.jmb.2018.05.030

260. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521,** 81–84 (2015).

261. Li, J. *et al.* TCPA: a resource for cancer functional proteomics data. *Nat Methods* **10,** 1046–1047 (2013).

262. Borcherding, N., Bormann, N. L., Voigt, A. P. & Zhang, W. TRGAted: A web tool for survival analysis using protein data in the Cancer Genome Atlas. *F1000Research* **7,** (2018).

263. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2,** 401–404 (2012).

264. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6,** pl1 (2013).

265. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57,** 289–300 (1995).

266. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4,** 2612 (2013).

267. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12,** 453 (2015).

268. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* (2014). doi:10.1038/nmeth.3047

269. Kolb, R. *et al.* Obesity-associated NLRC4 inflammasome activation drives breast cancer progression. *Nat. Commun.* (2016). doi:10.1038/ncomms13007

270. Edelbrock, M. A., Kaliyaperumal, S. & Williams, K. J. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **743–744,** 53–66 (2013).

271. Zhang, M. *et al.* HDAC6 deacetylates and ubiquitinates MSH2 to maintain proper levels of MutSα. *Mol. Cell* (2014). doi:10.1016/j.molcel.2014.04.028

272. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* (2016). doi:10.1038/nature18003

273. Vinci, M. *et al.* Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC Biol.* (2012). doi:10.1186/1741-7007-10-29

274. Shang, B., Liu, Y., Jiang, S. & Liu, Y. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. *Sci. Rep.* (2015). doi:10.1038/srep15179

275. Dawood, S., Broglio, K., Buzdar, A. U., Hortobagyi, G. N. & Giordano, S. H. Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: An institutional-based review. *J. Clin. Oncol.* **28,** 92–98 (2010).

276. Cristofanilli, M. *et al.* Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phas. *Lancet Oncol.* **17,** 425–439 (2016).

277. DiPippo, A. J., Patel, N. K. & Barnett, C. M. Cyclin-Dependent Kinase Inhibitors for the Treatment of Breast Cancer: Past, Present, and Future. *Pharmacotherapy* **36,** 652–667 (2016).

278. Garcï¿½a-Teijido, P., Cabal, M. L., Fernï¿½ndez, I. P. & Pï¿½rez, Y. F. Tumor-infiltrating lymphocytes in triple negative breast cancer: The future of immune targeting. *Clinical Medicine Insights: Oncology* **10,** 31–39 (2016).

279. Loi, S. *et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann. Oncol.* **25,** 1544–50 (2014).

280. Adams, S., Goldstein, L. J., Sparano, J. A., Demaria, S. & Badve, S. S. Tumor infiltrating lymphocytes (TILs) improve prognosis in patients with triple negative breast cancer (TNBC). *Oncoimmunology* **4,** 1–3 (2015).

281. Denkert, C. *et al.* Evaluation of tumorinfiltrating lymphocytes (TILs) as predictive and prognostic biomarker in different subtypes of breast cancer treated with neoadjuvant therapy-A metaanalysis of 3771 patients. *Cancer Res.* **77,** (2017).

282. Loi, S. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J. Clin. Oncol.* **31,** 860–867 (2013).

283. Dieci, M. V. *et al.* Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: A retrospective multicenter study. *Ann. Oncol.* **25,** 611–618 (2014).

284. Adams, S. *et al.* Phase 2 study of pembrolizumab (pembro) monotherapy for previously treated metastatic triple-negative breast cancer (mTNBC): KEYNOTE-086 cohort A. *J. Clin. Oncol.* **35,** Abstract 1008 (2017).

285. Dirix, L. *et al.* Avelumab (MSB0010718C), an anti-PD-L1 antibody, in patients with locally advanced or metastatic breast cancer: A phase Ib JAVELIN solid tumor trial. in *San Antonio Breast Cancer Symposium* S1-04 (2014). doi:10.1158/1538-7445.SABCS15-S1-04

286. Bauer, K. *et al.* T cell responses against microsatellite instability-induced frameshift peptides and influence of regulatory T cells in colorectal cancer. *Cancer Immunol. Immunother.* (2013). doi:10.1007/s00262-012-1303-8

287. Michel, S. *et al.* High density of FOXP3-positive T cells infiltrating colorectal cancers with microsatellite instability. *Br. J. Cancer* (2008). doi:10.1038/sj.bjc.6604756

288. Guidoboni, M. *et al.* Microsatellite instability and high content of activated cytotoxic lymphocytes identify colon cancer patients with a favorable prognosis. *Am. J. Pathol.* (2001). doi:10.1016/S0002-9440(10)61695-1

289. Network, T. C. G. A. *et al.* Comprehensive molecular portraits of human breast tumors. *Nature* **490,** 61–70 (2012).

290. Le Mercier, I., Lines, J. L. & Noelle, R. J. Beyond CTLA-4 and PD-1, the Generation Z of Negative Checkpoint Regulators. *Front Immunol* **6,** 418 (2015).

291. Sehrawat, S. *et al.* Galectin-9/TIM-3 interaction regulates virus-specific primary and memory CD8 T cell response. *PLoS Pathog* **6,** e1000882 (2010).

292. Konstantinopoulos, P. A. *et al.* TOPACIO/Keynote-162 (NCT02657889): A phase 1/2 study of niraparib + pembrolizumab in patients (pts) with advanced triple-negative breast cancer or recurrent ovarian cancer (ROC)—Results from ROC cohort. *J. Clin. Oncol.* **36,** 106 (2018).

293. Langer, C. J. *et al.* Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol.* **17,** 1497–1508 (2016).

294. Gandhi, L. *et al.* Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* NEJMoa1801005 (2018). doi:10.1056/NEJMoa1801005

295. Schmidt, A. M. *et al.* Regulatory T Cells Require TCR Signaling for Their Suppressive Function. *J. Immunol.* (2015). doi:10.4049/jimmunol.1402384

296. Tanaka, A. & Sakaguchi, S. Regulatory T cells in cancer immunotherapy. *Cell Research* (2017). doi:10.1038/cr.2016.151

297. Simpson, T. R. *et al.* Fc-dependent depletion of tumor-infiltrating regulatory T cells co-defines the efficacy of anti–CTLA-4 therapy against melanoma. *J. Exp. Med.* (2013). doi:10.1084/jem.20130579

298. Zou, W., Wolchok, J. D. & Chen, L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci. Transl. Med.* (2016). doi:10.1126/scitranslmed.aad7118

299. Arce Vargas, F. *et al.* Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. *Cancer Cell* **33,** 649–663.e4 (2018).

300. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* **4,** 330–336 (2003).

301. Hori, S., Nomura, T. & Sakaguchi, S. Control of Regulatory T Cell Development by the Transcription Factor. *Science (80-. ).* **299,** 1057 LP-1061 (2003).

302. Liu, S. *et al.* Prognostic significance of FOXP3+ tumor-infiltrating lymphocytes in breast cancer depends on estrogen receptor and human epidermal growth factor receptor-2 expression status and concurrent cytotoxic T-cell infiltration. *Breast Cancer Res.* (2014). doi:10.1186/s13058-014-0432-8

303. Sasaki, a *et al.* Prognostic value of tumor-infiltrating FOXP3+ regulatory T cells in patients with hepatocellular carcinoma. *Eur. J. Surg. Oncol.* (2008). doi:10.1016/j.ejso.2007.08.008

304. Miyara, M. *et al.* Functional Delineation and Differentiation Dynamics of Human CD4+T Cells Expressing the FoxP3 Transcription Factor. *Immunity* (2009). doi:10.1016/j.immuni.2009.03.019

305. Saito, T. *et al.* Two FOXP3+CD4+ T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat. Med.* (2016). doi:10.1038/nm.4086

306. Lowther, D. E. *et al.* PD-1 marks dysfunctional regulatory T cells in malignant gliomas. *JCI Insight* (2016). doi:10.1172/jci.insight.85935

307. De Simone, M. *et al.* Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. *Immunity* (2016). doi:10.1016/j.immuni.2016.10.021

308. Plitas, G. *et al.* Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. *Immunity* (2016). doi:10.1016/j.immuni.2016.10.032

309. Zheng, C. *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* (2017). doi:10.1016/j.cell.2017.05.035

310. Kissel, K., Santoso, S., Hofmann, C., Stroncek, D. & Bux, J. Molecular basis of the neutrophil glycoprotein NB1 (CD177) involved in the pathogenesis of immune neutropenias and transfusion reactions. *Eur. J. Immunol.* (2001). doi:10.1002/1521-4141(200105)31:5<1301::AID-IMMU1301>3.0.CO;2-J

311. Stroncek, D. F. Neutrophil-specific antigen HNA-2a, NB1 glycoprotein, and CD177. *Current Opinion in Hematology* (2007). doi:10.1097/MOH.0b013e3282efed9e

312. Xie, Q. *et al.* Characterization of a novel mouse model with genetic deletion of CD177. *Protein Cell* (2015). doi:10.1007/s13238-014-0109-1

313. Toyoda, T. *et al.* Gene expression analysis of a Helicobacter pylori-infected and high-salt diet-treated mouse gastric tumor model: identification of CD177 as a novel prognostic factor in patients with gastric cancer. *BMC Gastroenterol.* (2013). doi:10.1186/1471-230X-13-122

314. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* (2011). doi:10.1038/nbt.2038

315. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** (2017).

316. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).

317. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4096

318. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14,** 979–982 (2017).

319. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).

320. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* (2017). doi:10.1038/nmeth.4324

321. Biswas, T., Gu, X., Yang, J., Ellies, L. G. & Sun, L. Z. Attenuation of TGF-β signaling supports tumor progression of a mesenchymal-like mammary tumor cell line in a syngeneic murine model. *Cancer Lett.* (2014). doi:10.1016/j.canlet.2013.12.018

322. Iellem, A. *et al.* Unique chemotactic response profile and specific expression of chemokine receptors CCR4 and CCR8 by CD4(+)CD25(+) regulatory T cells. *J. Exp. Med.* (2001). doi:10.1084/jem.194.6.847

323. Sachs, U. J. H. *et al.* The neutrophil-specific antigen CD177 is a counter-receptor for platelet endothelial cell adhesion molecule-1 (CD31). *J. Biol. Chem.* (2007). doi:10.1074/jbc.M701120200

324. Willemze, R. *et al.* WHO-EORTC classification for cutaneous lymphomas. *Blood* **105,** 3768–3785 (2005).

325. Kirsch, I. R. *et al.* TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci. Transl. Med.* **7,** (2015).

326. Clark, R. A. *et al.* High-scatter T cells: A reliable biomarker for malignant T cells in cutaneous T-cell lymphoma. *Blood* **117,** 1966–1976 (2011).

327. Ormsby, A., Bergfeld, W. F., Tubbs, R. R. & Hsi, E. D. Evaluation of a new paraffin-reactive CD7 T-cell deletion marker and a polymerase chain reaction-based T-cell receptor gene rearrangement assay: Implications for diagnosis of mycosis fungoides in community clinical practice. *J. Am. Acad. Dermatol.* **45,** 405–413 (2001).

328. Wilcox, R. A. Cutaneous T-cell lymphoma: 2017 update on diagnosis, risk-stratification, and management. *Am. J. Hematol.* **92,** 1085–1102 (2017).

329. Michie, S. A., Abel, E. A., Hoppe, R. T., Warnke, R. A. & Wood, G. S. Discordant expression of antigens between intraepidermal and intradermal T cells in mycosis fungoides. *Am. J. Pathol.* **137,** 1447–51 (1990).

330. Agar, N. S. *et al.* Survival Outcomes and Prognostic Factors in Mycosis Fungoides/Sezary Syndrome: Validation of the Revised International Society for Cutaneous Lymphomas/European Organisation for Research and Treatment of Cancer Staging Proposal. *J. Clin. Oncol.* **28,** 4730–4739 (2010).

331. Shin, J. *et al.* Lesional gene expression profiling in cutaneous T-cell lymphoma reveals natural clusters associated with disease outcome. *Blood* **110,** 3015–3027 (2007).

332. Booken, N. *et al.* Sézary syndrome is a unique cutaneous T-cell lymphoma as identified by an expanded gene signature including diagnostic marker molecules CDO1 and DNM3. *Leukemia* **22,** 393–399 (2008).

333. Lee, C. S. *et al.* Transcriptome sequencing in Sezary syndrome identifies Sezary cell and mycosis fungoides-associated lncRNAs and novel transcripts. *Blood* **120,** 3288–3297 (2012).

334. van Doorn, R. *et al.* Oncogenomic analysis of mycosis fungoides reveals major differences with Sezary syndrome. *Blood* **113,** 127–36 (2009).

335. Choi, J. *et al.* Genomic landscape of cutaneous T cell lymphoma. *Nat. Genet.* **47,** 1011–1019 (2015).

336. Berger, C. L. *et al.* Cutaneous T-cell lymphoma: Malignant proliferation of T-regulatory cells. *Blood* **105,** 1640–1647 (2005).

337. Heid, J. B. *et al.* FOXP3+CD25- tumor cells with regulatory function in Sezary syndrome. *J Invest Dermatol* **129,** 2875–2885 (2009).

338. Tiemessen, M. M. *et al.* Lack of suppressive CD4+CD25+FOXP3+ T cells in advanced stages of primary cutaneous T-cell lymphoma. *J. Invest. Dermatol.* **126,** 2217–23 (2006).

339. Krejsgaard, T. *et al.* Malignant Tregs express low molecular splice forms of FOXP3 in Sézary syndrome. *Leukemia* **22,** 2230–2239 (2008).

340. Gjerdrum, L. M. *et al.* FOXP3+ regulatory T cells in cutaneous T-cell lymphomas: association with disease stage and survival. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K* **21,** 2512–2518 (2007).

341. Buus, T. B. *et al.* Single-cell heterogeneity in Sézary syndrome. *Blood Adv.* **2,** 2115–2126 (2018).

342. Aran, D. *et al.* Reference-based annotation of single-cell transcriptomes identifies a profibrotic macrophage niche after tissue injury. *bioRxiv* 284604 (2018). doi:10.1101/284604

343. Lefrançois, P. *et al.* TruSeq-Based Gene Expression Analysis of Formalin-Fixed Paraffin-Embedded (FFPE) Cutaneous T-Cell Lymphoma Samples: Subgroup Analysis Results and Elucidation of Biases from FFPE Sample Processing on the TruSeq Platform. *Front. Med.* **4,** (2017).

344. Ferenczi, K., Fuhlbrigge, R. C., Pinkus, J. L., Pinkus, G. S. & Kupper, T. S. Increased CCR4 expression in cutaneous T cell lymphoma. *J. Invest. Dermatol.* **119,** 1405–1410 (2002).

345. Kari, L. *et al.* Classification and Prediction of Survival in Patients with the Leukemic Phase of Cutaneous T Cell Lymphoma. *J. Exp. Med.* **197,** 1477–1488 (2003).

346. Nebozhyn, M. *et al.* Quantitative PCR on 5 genes reliably identifies CTCL patients with 5% to 99% circulating tumor cells with 90% accuracy. *Blood* **107,** 3189–3196 (2006).

347. Wang, Y. *et al.* Deficiency of SATB1 expression in Sézary cells causes apoptosis resistance by regulating FasL/CD95L transcription. *Blood* **117,** 3826–3835 (2011).

348. Mao, X. *et al.* Amplification and overexpression of JUNB is associated with primary cutaneous T-cell lymphomas. *Blood* **101,** 1513–1519 (2003).

349. Poszepczynska-Guigné, E. *et al.* CD158k/KIR3DL2 is a new phenotypic marker of sezary cells: Relevance for the diagnosis and follow-up of sezary syndrome. *J. Invest. Dermatol.* **122,** 820–823 (2004).

350. Su, M. W. *et al.* Aberrant Expression of T-Plastin in Sezary Cells. *Cancer Res.* **63,** 7122–7127 (2003).

351. Zhang, Y. *et al.* Molecular markers of early-stage mycosis fungoides. *J. Invest. Dermatol.* **132,** 1698–1706 (2012).

352. Samimi, S. *et al.* Increased Programmed Death-1 Expression on CD4+ T Cells in Cutaneous T-Cell Lymphoma. *Arch. Dermatol.* **146,** 1382 (2010).

353. Dulmage BO, Akilov O, Vu JR, Falo LD, G. L. Dysregulation of the TOX-RUNX3 pathway in cutaneous T-cell lymphoma. *Oncotarget* **Nov 6,** (2015).

354. Huang, Y., Su, M. W., Jiang, X. & Zhou, Y. Evidence of an oncogenic role of aberrant TOX activation in cutaneous T-cell lymphoma. *Blood* **125,** 1435–1443 (2015).

355. Dudoit, S., Fridlyand, J. & Speed, T. P. Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data Comparison of Discrimination Methods for the Classi' cation of Tumors Using Gene Expression Data. *J. Am. Stat. Assoc.* **97457,** 77–87 (2002).

356. Hori, S., Nomura, T. & Sakaguchi, S. Control of Regulatory T Cell Development by the Transcription Factor. *Science (80-. ).* **299,** 1057 LP-1061 (2003).

357. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* **4,** 330–336 (2003).

358. Komatsu, N. *et al.* Heterogeneity of natural Foxp3 $^{+}$ T cells: A committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl. Acad. Sci.* **106,** 1903–1908 (2009).

359. Williams, L. M. & Rudensky, A. Y. Maintenance of the Foxp3-dependent developmental program in mature regulatory T cells requires continued expression of Foxp3. *Nat. Immunol.* **8,** 277–284 (2007).

360. Wan, Y. Y. & Flavell, R. A. Regulatory T-cell functions are subverted and converted owing to attenuated Foxp3 expression. *Nature* **445,** 766–770 (2007).

361. Tao, R. *et al.* Deacetylase inhibition promotes the generation and function of regulatory T cells. *Nat. Med.* **13,** 1299–1307 (2007).

362. Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R. & Pazdur, R. FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **12,** 1247–52 (2007).

363. Satpathy, A. T. *et al.* Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **24,** 580–590 (2018).

364. Maroulakou, I. G., Anver, M., Garrett, L. & Green, J. E. Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. *Proc. Natl. Acad. Sci. U. S. A.* (1994). doi:10.1073/pnas.91.23.11236

365. Ozawa, M. & Kemler, R. Molecular organization of the uvomorulin-catenin complex. *J. Cell Biol.* **116,** 989–996 (1992).

366. Birchmeier, W., Jauch, K. W., Heiss, M. M., Schildberg, F. W. & Funke, I. E-Cadherin Expression in Primary and Metastatic Gastric Cancer: Down-Regulation Correlates with Cellular Dedifferentiation and Glandular Disintegration. *Cancer Res.* **53,** 1690–1695 (1993).

367. Dorudi, S., Sheffield, J. P., Poulsom, R., Northover, J. M. A. & Hart, I. R. E-cadherin expression in colorectal cancer: An immunocytochemical and in situ hybridization study. *Am. J. Pathol.* **142,** 981–986 (1993).

368. Bremnes, R. M., Veve, R., Hirsch, F. R. & Franklin, W. A. The E-cadherin cell-cell adhesion complex and lung cancer invasion, metastasis, and prognosis. *Lung Cancer* **36,** 115–124 (2002).

369. Bringuier, P. P. *et al.* Decreased E-cadherin immunoreactivity correlates with poor survival in patients with bladder tumors. *Cancer Res* **53,** 3241–3245 (1993).

370. Umbas, R. *et al.* Decreased E-cadherin expression is associated with poor prognosis in patients with prostate cancer. *Cancer Res.* **54,** 3929–3933 (1994).

371. Oka, H. *et al.* Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis. *Cancer Res.* **53,** 1696–1701 (1993).

372. Takeichi, M. Cadherins in cancer: implications for invasion and metastasis. *Curr. Opin. Cell Biol.* **5,** 806–811 (1993).

373. Berx, G. & Van Roy, F. The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. *Breast Cancer Res.* **3,** 289 (2001).

374. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158,** 1110–1122 (2014).

375. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434,** 843 (2005).

376. Aberle, H. *et al.* Assembly of the cadherin-catenin complex in vitro with recombinant proteins. *J. Cell Sci.* **107,** 3655–3663 (1994).

377. Hinck, L., Näthke, I. S., Papkoff, J. & Nelson, W. J. Dynamics of cadherin/catenin complex formation: novel protein interactions and pathways of complex assembly. *J. Cell Biol.* **125,** 1327–1340 (1994).

378. Nelson, W. J. & Nusse, R. Convergence of Wnt, ß-catenin, and cadherin pathways. *Science (80-. ).* **303,** 1483–1487 (2004).

379. Yoshida, R., Kimura, N., Harada, Y. & Ohuchi, N. The loss of E-cadherin, α-and β-catenin expression is associated with metastasis and poor prognosis in invasive breast cancer. *Int. J. Oncol.* **18,** 513–520 (2001).

380. Zschiesche, W. *et al.* Expression of E-cadherin and catenins in invasive mammary carcinomas. *Anticancer Res.* **17,** 561–567 (1997).

381. Dillon, D. A., D'Aquila, T., Reynolds, A. B., Fearon, E. R. & Rimm, D. L. The expression of p120ctn protein in breast cancer is independent of alpha-and beta-catenin and E-cadherin. *Am. J. Pathol.* **152,** 75 (1998).

382. Brzozowska, A., Sodolski, T., Duma, D., Mazurkiewicz, T. & Mazurkiewicz, M. Evaluation of prognostic parameters of E-cadherin status in breast cancer treatment. *Ann. Agric. Environ. Med.* **19,** (2012).

383. Nakagawa, M. *et al.* Expression of p53, Ki-67, E-cadherin, N-cadherin and TOP2A in triple-negative breast cancer. *Anticancer Res.* **31,** 2389–2393 (2011).

384. Siitonen, S. M. *et al.* Reduced E-cadherin expression is associated with invasiveness and unfavorable prognosis in breast cancer. *Am. J. Clin. Pathol.* **105,** 394–402 (1996).

385. Gillett, C. E. *et al.* Retention of the expression of E-cadherin and catenins is associated with shorter survival in grade III ductal carcinoma of the breast. *J. Pathol.* **193,** 433–441 (2001).

386. Li, C. I., Anderson, B. O., Daling, J. R. & Moe, R. E. Trends in incidence rates of invasive lobular and ductal breast carcinoma. *Jama* **289,** 1421–1424 (2003).

387. Arpino, G., Bardou, V. J., Clark, G. M. & Elledge, R. M. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast cancer Res.* **6,** R149 (2004).

388. Toikkanen, S., Pylkkänen, L. & Joensuu, H. Invasive lobular carcinoma of the breast has better short- and long-term survival than invasive ductal carcinoma. *Br. J. Cancer* **76,** 1234 (1997).

389. Adachi, Y. *et al.* Comparison of clinical outcomes between luminal invasive ductal carcinoma and luminal invasive lobular carcinoma. *BMC Cancer* **16,** 248 (2016).

390. Li, C. I., Moe, R. E. & Daling, J. R. Risk of mortality by histologic type of breast cancer among women aged 50 to 79 years. *Arch. Intern. Med.* **163,** 2149–2153 (2003).

391. Chen, Z. *et al.* Invasive lobular carcinoma of the breast: A special histological type compared with invasive ductal carcinoma. *PLoS One* **12,** e0182397 (2017).

392. Ciriello, G. *et al.* Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163,** 506–519 (2015).

393. de Leeuw, W. J. F. *et al.* Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J. Pathol. A J. Pathol. Soc. Gt. Britain Irel.* **183,** 404–411 (1997).

394. Sarrió, D. *et al.* Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability. *Int. J. cancer* **106,** 208–215 (2003).

395. Kowalski, P. J., Rubin, M. A. & Kleer, C. G. E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast Cancer Res.* **5,** R217 (2003).

396. Gonzalez, M. A. *et al.* An immunohistochemical examination of the expression of E-cadherin, α-and β/γ-catenins, and α2-and β1-integrins in invasive breast cancer. *J. Pathol.* **187,** 523–529 (1999).

397. Bukholm, I. K., Nesland, J. M., Kåresen, R., Jacobsen, U. & Børresen-Dale, A. E-cadherin and α-, β-, and γ-catenin protein expression in relation to metastasis in human breast carcinoma. *J. Pathol. A J. Pathol. Soc. Gt. Britain Irel.* **185,** 262–266 (1998).

398. Li, J. *et al.* TCPA: a resource for cancer functional proteomics data. *Nat Methods* **10,** 1046–1047 (2013).

399. Cline, M. S. *et al.* Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. *Sci Rep* **3,** 2652 (2013).

400. Győrffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123,** 725–731 (2010).

401. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346 (2012).

402. Dabbs, D. J., Bhargava, R. & Chivukula, M. Lobular versus ductal breast neoplasms: the diagnostic utility of p120 catenin. *Am. J. Surg. Pathol.* **31,** 427–437 (2007).

403. ElMoneim, H. M. A. & Zaghloul, N. M. Expression of E-cadherin, N-cadherin and snail and their correlation with clinicopathological variants: an immunohistochemical study of 132 invasive ductal breast carcinomas in Egypt. *Clinics* **66,** 1765–1771 (2011).

404. Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123,** 725–731 (2010).

405. Geyer, F. C. *et al.* β-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod. Pathol.* **24,** 209 (2011).

406. Tan, P. H. *et al.* Renal tumors: diagnostic and prognostic biomarkers. *Am. J. Surg. Pathol.* **37,** 1518 (2013).

407. Katagiri, A., Watanabe, R. & Tomita, Y. E-cadherin expression in renal cell cancer and its significance in metastasis and survival. *Br. J. Cancer* **71,** 376 (1995).

408. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science (80-. ).* **357,** eaan2507 (2017).

409. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (80-. ).* **347,** 1260419 (2015).

410. Xing, X. *et al.* The prognostic value of E-cadherin in gastric cancer: A meta-analysis. *Int. J. cancer* **132,** 2589–2596 (2013).

411. Chen, J. *et al.* Prognostic significance of E-cadherin expression in hepatocellular carcinoma: a meta-analysis. *PLoS One* **9,** e103952 (2014).

412. Xie, Y. *et al.* Reduced E-cadherin expression is correlated with poor prognosis in patients with bladder cancer: a systematic review and meta-analysis. *Oncotarget* **8,** 62489 (2017).

413. Li, Z., Yin, S., Zhang, L., Liu, W. & Chen, B. Prognostic value of reduced E-cadherin expression in breast cancer: a meta-analysis. *Oncotarget* **8,** 16445 (2017).

414. He, X., Chen, Z., Jia, M. & Zhao, X. Downregulated E-cadherin expression indicates worse prognosis in Asian patients with colorectal cancer: evidence from meta-analysis. *PLoS One* **8,** e70858 (2013).

415. Rothberg, B. E. G. & Bracken, M. B. E-cadherin immunohistochemical expression as a prognostic factor in infiltrating ductal carcinoma of the breast: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **100,** 139–148 (2006).

416. Tibes, R. *et al.* Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol. Cancer Ther.* **5,** 2512–2521 (2006).

417. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* **110,** 4009–4014 (2013).

418. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6,** 8971 (2015).

419. Zheng, X., Zhang, N., Wu, H.-J. & Wu, H. Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome Biol.* **18,** 17 (2017).

420. Rakha, E. A. *et al.* Prognostic markers in triple-negative breast cancer. *Cancer* **109,** 25–32 (2007).

421. Shamir, E. R. & Ewald, A. J. Adhesion in mammary development: novel roles for E-cadherin in individual and collective cell migration. in *Current topics in developmental biology* **112,** 353–382 (Elsevier, 2015).

422. Nieman, M. T., Prudoff, R. S., Johnson, K. R. & Wheelock, M. J. N-cadherin promotes motility in human breast cancer cells regardless of their E-cadherin expression. *J. Cell Biol.* **147,** 631–644 (1999).

423. López-Knowles, E. *et al.* Cytoplasmic localization of β-catenin is a marker of poor outcome in breast cancer patients. *Cancer Epidemiol. Prev. Biomarkers* **19,** 301–309 (2010).

424. Dolled-Filhart, M. *et al.* Quantitative in situ analysis of β-catenin expression in breast cancer shows decreased expression is associated with poor outcome. *Cancer Res.* **66,** 5487–5494 (2006).

425. Lin, S.-Y. *et al.* β-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. *Proc. Natl. Acad. Sci.* **97,** 4262–4266 (2000).

426. Klemm, F. *et al.* β-catenin-independent WNT signaling in basal-like breast cancer and brain metastasis. *Carcinogenesis* **32,** 434–442 (2010).

427. Khramtsov, A. I. *et al.* Wnt/β-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am. J. Pathol.* **176,** 2911–2920 (2010).

428. Bui, T. *et al.* β-Catenin haploinsufficiency promotes mammary tumorigenesis in an ErbB2-positive basal breast cancer model. *Proc. Natl. Acad. Sci.* **114,** E707–E716 (2017).

429. Xi, X. *et al.* RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Noncoding RNA* **3,** (2017).

430. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2,** 401–404 (2012).

431. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6,** pl1 (2013).

432. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173,** 400–416.e11 (2018).

433. Wright, M. N., Dankowski, T. & Ziegler, A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Stat Med* **36,** 1272–1284 (2017).