UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVII CICLO DEL DOTTORATO DI RICERCA IN

MOLECULAR BIOMEDICINE

Understanding the complexity of the tumor microenvironment through an integrated bioinformatics approach

Settore scientifico-disciplinare: BIO/13

DOTTORANDO
LUCA TRIBOLI

COORDINATORE
PROF. ALESSANDRO TOSSI

SUPERVISORE DI TESI
PROF. GIANNINO DEL SAL

CO-SUPERVISORE DI TESI
DR. SILVANO PIAZZA

ANNO ACCADEMICO 2023/2024

INDEX

ABSTRACT iv

1. INTRODUCTION 1

1.1. Tumor Heterogeneity Impacts Therapy Outcome 1

1.2 Impact of the TME on Tumor Development and Progression 2

1.3 Impact of the TME on Anti-Cancer Therapies 6

1.4 Impact of Oncogenic Signaling on the Evolution of the TME 8

1.5 The Prolyl-Isomerase PIN1: a Central Hub in Oncogenic Signaling and Therapy Resistance 10

1.6 Bioinformatics Approaches for Precision Medicine 12

2.AIM OF THE THESIS 17

3.RESULTS 19

3.1 Download Genome Sequences, Gene and TEs Annotations 20

3.2 Set Up an Automatic Alignment Pipeline 23

3.3 Generation of Molecular and Phenotypic Signatures Scores from RNAseq Experiments . 25

3.4 Annotation and Signatures Correlation of BC Patients 31

3.5 PIN1 Knockdown Reactivates an Innate Immune Response and Derepresses Heterochromatin 34

3.6 Ablation of PIN1 in the 4T1 Mouse Model Generates Immunogenic Neoantigens 37

3.7 scRNAseq from cPIN1KO in PyMT Mouse Model Shows Increased Immune Infiltrate and Activation of Both Innate and Adaptive Immune Response 43

4. MATERIAL AND METHODS 47

5. DISCUSSION AND FUTURE DIRECTIONS 49

5.1 Relationship of PIN1 activity with ECM stiffness and the immune TME 49

5.2 Therapeutic opportunities 50

5.3 The Challenge of Breast Cancers Subtypes 50

5.4 Advances and limitations in sequencing resolution for the analysis of the impact of PIN1 in the TME 51

5.5 Artificial Intelligence (AI) and Omics approaches in cancer research 52

6. ACKNOWLEDGEMENTS 53

REFERENCES 55

ABSTRACT

Cancer stands as a prominent contributor to global mortality, accounting for 20 million new cancer cases and 9.7 million deaths globally in 2022, according to the latest Report from the International Agency for Research (IARC). Cancer is a disease of uncontrolled proliferation of transformed cells that accumulate genetic and epigenetic alterations, which endow these cells with new biological properties driving tumor initiation, progression, and treatment resistance. Yet, although cancer cell-intrinsic factors influence treatment efficacy, significant contribution to tumor growth and response to therapy is provided by the tumor microenvironment (TME), which includes various stromal (e.g. endothelial cells and fibroblasts) and immune cells, and the non-cellular components of the extracellular matrix (ECM). Within the TME, the dynamic interactions between cancer and non-cancer cells and the ECM, which occur through chemical and mechanical cues, are critical determinants for progression to malignant and treatment resistant stages.

The TME can have either tumor-suppressing or tumor-promoting functions. Aggressive cancers are often characterized by heterogeneous and dynamic TMEs, which evolve immune-suppressive mechanisms allowing evasion from anti-tumor immune reactions. Yet, the understanding of the molecular mechanisms driving the behaviour of the tumor as an ecosystem is still quite limited and could be improved through approaches that integrate the study of tumor cells and TME.

Along this direction, a transformative change for understanding the complexity of tumor heterogeneity is provided by bioinformatic studies that integrate clinical characteristics and biological omics data obtained through next-generation sequencing (NGS) approaches. Traditionally, cancer research relied on bulk RNA sequencing, which provides an aggregate gene expression profile from a tumor sample. However, this approach masks the intricate cellular diversity within tumors. More advanced technologies, like single cell transcriptomic and in situ spatial transcriptomics, were used on tumors and their TME to identify cell types, distinguish gene expression of different cell lineages, as well as to extract information on the crosstalk between neighboring cells, thus representing powerful tools to unravel tumor heterogeneity and molecular mechanisms underlying drug response. This multi-dimensional view represents a paradigm shift in cancer research, moving from a simplified, averaged perspective to a nuanced understanding of tumors as complex, dynamic ecosystems with intricate cellular interactions. The results from

integrated data will also help reconsidering standard tumor classification methods that mainly rely on cancer cell-intrinsic features.

In this work, to identify molecular mechanisms and determinants that could modulate the TME and drive therapy resistance, we focused on a particular enzyme called PIN1, a prolyl-isomerase that regulates multiple cancer-driving pathways by controlling the conformational change of many oncogenic proteins following phosphorylation by proline directed kinases (CDKs and MAPKs). PIN1 is prevalently overexpressed in human tumors and linked to therapy resistance, and in cell and animal models of breast cancer (BC) it has been demonstrated that genetic or pharmacological inhibition of PIN1 dramatically curbs tumor growth, metastatic progression and treatment resistance. BC is generally considered an immune "cold" tumor, mainly due to poor immune cell infiltration or to presence of inactivated immune cells. An emerging feature of such TMEs is the high stiffness of the ECM, and enhanced cellular mechano-signaling, which drives tumor development and therapy resistance, including the generation of an immune-suppressive TME.

Thus, to understand whether and how PIN1 could be involved in reshaping the TME in BC, we investigated the correlation between PIN1 activity and features of BC aggressiveness, such as ECM stiffness and immune cell status, using transcriptional signatures as surrogate. Analysis of BC transcriptomic data with these signatures showed a direct correlation between PIN1 activity and ECM stiffness. Conversely, both PIN1 activity and ECM stiffness inversely correlated with the presence of anti-tumor immune cells. In parallel, we investigated whether and how the ablation of PIN1 in cancer cells impact the TME, both in of human BC cells treated with PIN1 siRNA, and two mouse models of immune cold BC, in which PIN1 could be knocked out (KO) or knocked down (KD) in an inducible fashion (*MMTV-PyMT;R26CreERT2;Pin1*$^{fl/fl}$ and 4T1-TetO-*shPin1* allografts). Importantly, in these two in vivo models, PIN1KO/KD caused a robust reduction of tumor/metastasis growth that was accompanied by remarkable changes within the TME, as demonstrated by combining histological analyses with bulk and single cell RNAseq analyses. Integrated analysis of these data using recently developed bioinformatic pipelines showed that PIN1 loss triggered a tumor cell-intrinsic increase in both innate and adaptive immune pathways, as well as upregulation of transposable elements (TEs), a potential source of cytoplasmic nucleic acids triggering the cGAS/STING/IFN-I innate immunity pathway, and generation of TE-derived immunogenic antigens. These effects correlated, in the TME, with an increase of infiltrated and

activated immune cells of the innate (Natural Killer) and the adaptive (CD8+ T-cell) immune system.

Altogether these results suggest that PIN1 activity is supported by a stiff ECM and maintains an immunosuppressive TME in BC. Thus, PIN1 activity could improve classification of hot versus cold tumors, adding a missing layer towards the understanding of response to immune therapies. Furthermore, PIN1 inhibition with available FDA approved drugs may reduce tumor growth by reprogramming tumor cells and the TME and reawaken the immune system, thus improving therapy response especially in combination with immune checkpoint inhibitors.

1. INTRODUCTION

Cancer is a complex disease that arises due to a series of genetic and epigenetic alterations within cells that cause a growth advantage and unscheduled proliferation, thus evading tissue growth control mechanisms (Sanchez-Vega et al.; Sever and Brugge). Cancer development is driven by variations in the genomic, epigenomic, transcriptomic, and proteomic features of cancer cells, such as those caused by mutations in key genes that regulate critical cellular processes. These genetic changes can be inherited or acquired due to various factors such as environmental exposures, lifestyle, or mutations during DNA replication and can impact crucial tumor properties including proliferation, resistance to apoptosis, metastasis, metabolism and response to treatment.

Yet, tumors are composed not only of tumor cells, but also non-tumor cells, like stromal and immune cells, embedded within a surrounding extracellular matrix (ECM), together constituting the tumor microenvironment (TME). The TME plays a crucial role in shaping tumor growth, invasion into surrounding tissues, metastasis, and response to treatments. The specific composition of the TME of the primary site or of the metastases and the interactions between tumor cells and their microenvironment are significant factors that can contribute to tumor progression and lead to unfavorable clinical outcomes.

1.1. Tumor Heterogeneity Impacts Therapy Outcome

The complexity of cancer is amplified by its heterogeneity that could be viewed at different levels: genetic, epigenetic, transcriptomic and proteomic. Moreover, this heterogeneity exists at multiple scales within cancer, given by: differences among patients, even if the tumors originate from the same tissue type (inter-tumor heterogeneity); or within a single tumor, constituted by subclones each characterized by its own sub-TME (intra-tumor heterogeneity) (Grünwald et al.); metastases in different organs may develop distinct properties (inter-metastatic heterogeneity).

Among several types of cancers, Breast Cancers (BCs) are a paradigmatic example of heterogeneous cancers, whose treatment management is difficult. BC patients are currently classified in 5 major subtypes (Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like) (Curtis et al.) that should receive therapies according to this classification (for example anti-

HER2 agents). However, BC patients diagnosed with the same subtype often do not respond in the same way to the same therapy (Guo et al.) (Figure 1).

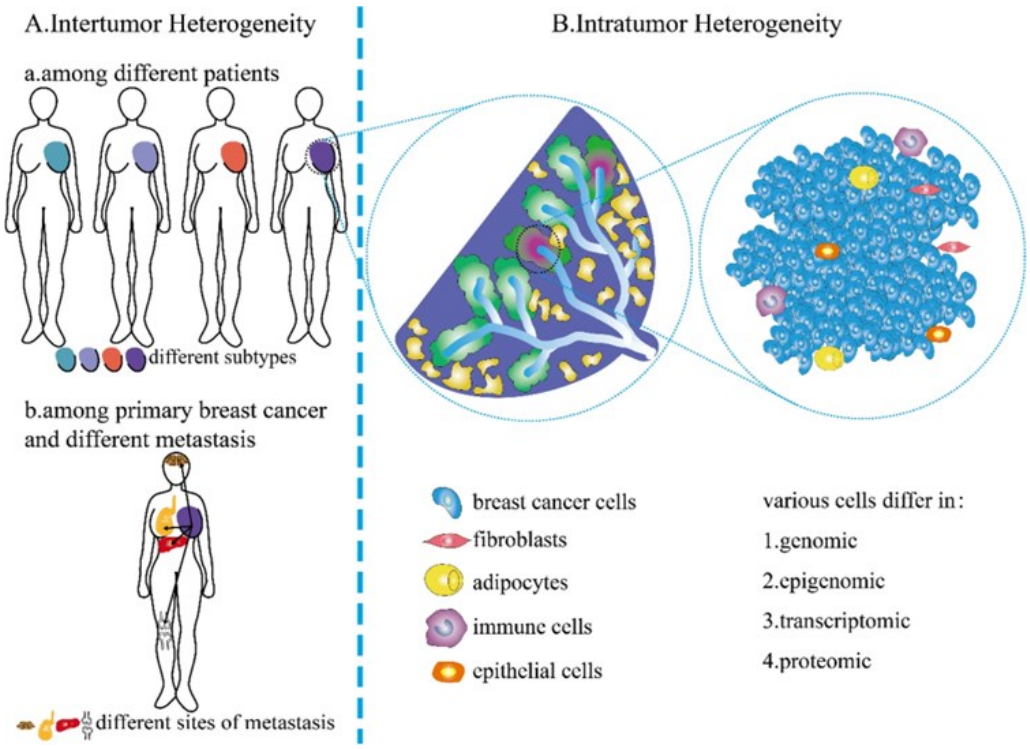

Figure 1: *Inter-tumor and intra-tumor heterogeneity.* Example of tumor heterogeneity in BC: among patients, within the same patient across different primary and metastatic niches and, within the same niche, among different clones and cell types (Guo et al.).

1.2 Impact of the TME on Tumor Development and Progression

The TME is highly dynamic and evolves with the tumor through extensive cell-cell and cell-ECM interactions, including exchange of soluble molecules (Emon et al.) (Figure 2): at the early stages of tumor development, the TME is more hostile with more abundant cytotoxic T lymphocytes, NK (natural killer) cells, or activated M1 macrophages, responsible for an active immune surveillance (Deepak et al.). As tumors progress to malignant stages, they develop mechanisms to suppress anti-tumor immunity and promote a supportive TME, producing growth factors to promote survival of cancer cells, increasing angiogenesis, gaining a migratory and invasive phenotype (EMT, epithelial-mesenchymal transition). Yet also interactions between tumor and host cells and with the ECM participate in tumor progression, i.e. cancer cell directly contribute to TME physical changes by overgrowing and modifying ECM composition, and indirectly cancer cells regulate the

activity of stromal cells, such as CAFs and macrophages and increase the overall mechanical stress (Maller et al.; Humphrey et al.; Jiang et al.; Northey et al.; Mohammadi and Sahai; Piersma et al.; Stowers et al.; Bertolio et al.; Hayward et al.).

Mechanical cues have been shown to confer transformed cells with cancer stem cell properties, enable evasion of the immune system, and induce proliferation and resistance to therapies (Piersma et al.; Panciera et al.). It is worth mentioning that BC tissue exhibits a significant increase in stiffness compared to normal breast tissue. Measurements using AFM (atomic force microscopy) indicate that BC tissue can be up to 10 times stiffer, with values around 5 kPa, whereas normal breast tissue typically has a stiffness of approximately 0.4 kPa (Pratt et al.).

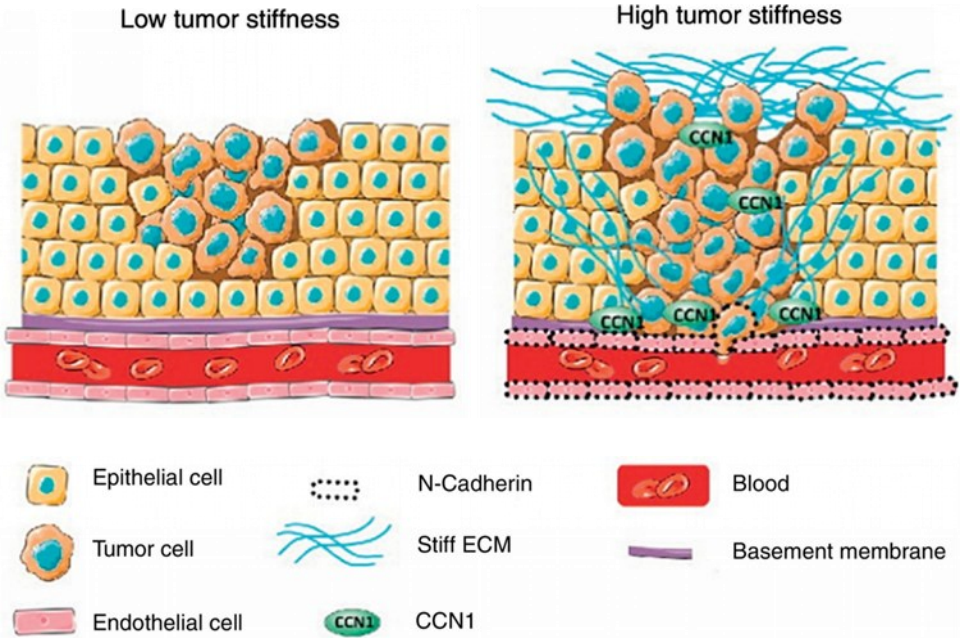

Figure 2: *Composition of tissue in different stiffness conditions.* When there is a stiffer extracellular matrix, tumor cells are more prone to tissue invasion and to enter the blood flow to migrate (Emon et al.).

The stiffness of the ECM also contributes to an increased expression of the Programmed Cell Death Ligand 1 (PD-L1), an immune checkpoint protein present on the surface of tumor cells, aiding in their evasion from immune surveillance. Moreover, elevated collagen density and ECM stiffness have been observed to decrease the abundance and effectiveness of cytotoxic T-cells within tumors (Nicolas-Boluda et al.; Kuczek et al.), while promoting the functional change of macrophages toward an immunosuppressive phenotype (Larsen et al.). There is also evidence that

mechanotransduction pathways deregulate innate immune response in tumor cells (Liu et al.), thus indicating that a stiff ECM not only promotes tumor cell growth and survival but also impacts the interaction between cancer cells and components of the TME. In fact, it was shown that in a TME characterized by high stiffness, activation of downstream mechanoresponsive pathways caused degradation of Cyclic GMP–AMP synthase (cGAS), a relevant protein involved in activation of innate immunity in response to presence of cytoplasmic DNA (Figure 3).

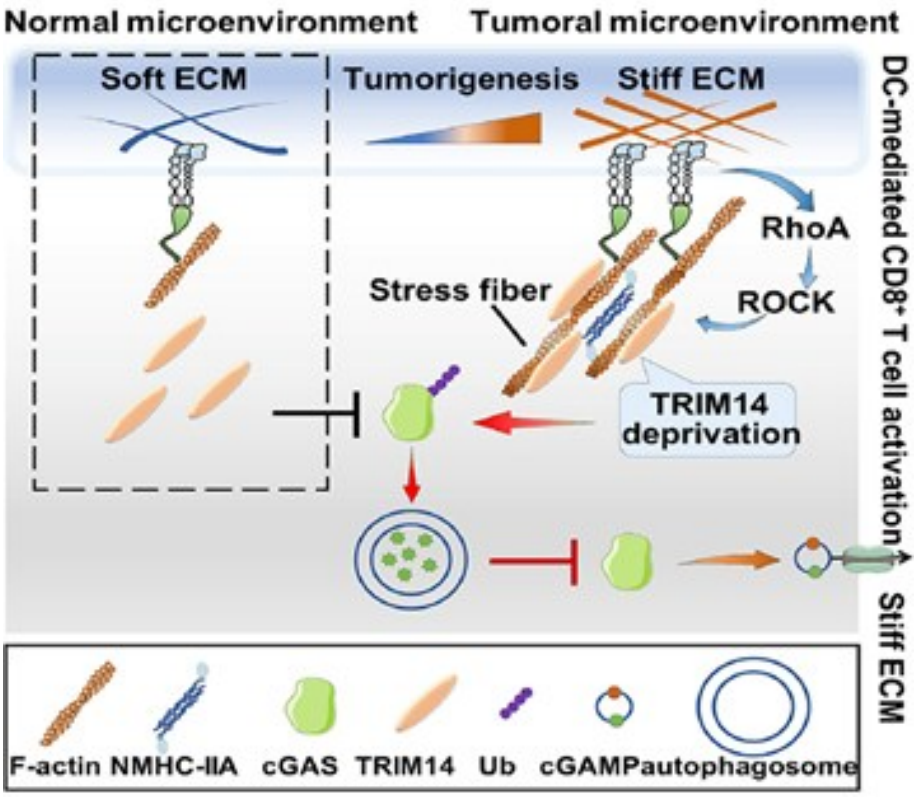

Figure 3: *Stiff ECM contributes to the immunosuppression in TME via mechano-signaling-dependent cGAS degradation*. The stiff ECM activates the ROCK (Rho-associated coiled-coil-containing protein kinase)-myosin IIA-actin pathway in tumor cells to promote the autophagic degradation of cGAS and dampen tumor antigenicity (Liu et al.).

cGAS activates STING (stimulator of interferon response cGAMP interactor) and this pathway was recently found to be one of the main pathways involved in innate immunity. It can sense and bind both endogenous and exogenous cytoplasmic dsDNA (double-stranded DNA), activating a downstream type I interferon response, which can promote recruitment of immune cells of the innate immune system. For example, exogenous DNA sources are virus infections while endogenous cytoplasmic dsDNA can derive from mitochondria or nuclei. Cytoplasmic DNA often derives from micronuclei, expression of genetic mobile sequences (Transposable Elements, TEs),

and once cytoplasmic DNA is sensed by cGAS, it oligomerizes with it in a 2:2 complex. This interaction activates cGAS that generates the second messenger 2',3'-cGAMP (Cyclic guanosine monophosphate–adenosine monophosphate). 2',3'-cGAMP stimulates STING at the Endoplasmic Reticulum (ER), then STING undergoes tetramerization and translocates from the ER to the Golgi and is palmitoylated. There, STING is phosphorylated by TBK1 (TANK-binding kinase 1) and recruits IRF3 (interferon regulatory factor 3) which in turn also undergoes phosphorylation by TBK1-mediated phosphorylation. Activated IRF3 dimerizes and translocates to the nucleus, where it promotes the transcription of Type I IFN genes which activate the innate immune response (Zierhut and Funabiki; Kwon and Bakhoum) (Figure 4).

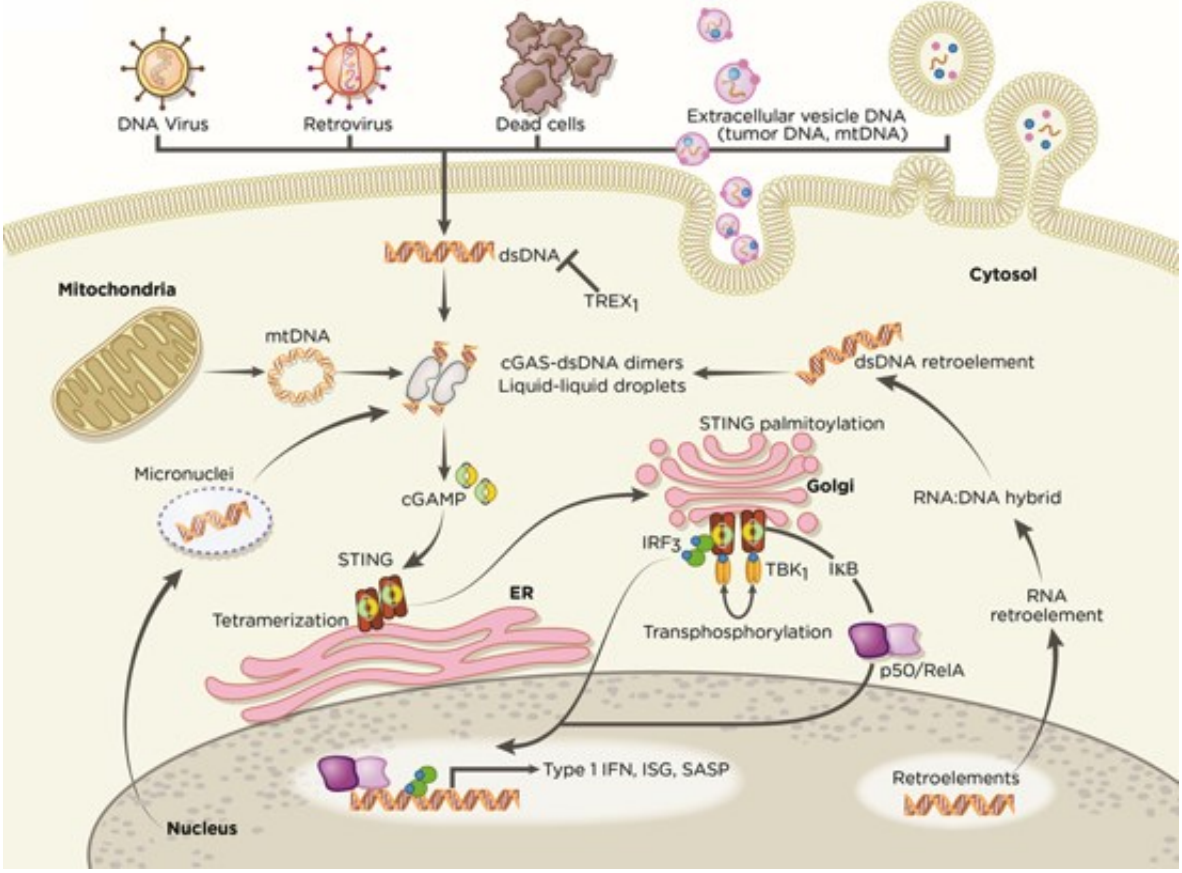

Figure 4: *Overview of the cGAS/STING1 pathway*. DsDNA coming from different sources are sensed by the cGAS molecule, which generates the second messenger cGAMP. It stimulates STING1, whose tetramerization activate TBK1 and IRF3, which promote the transcriptional expression of interferon genes (Kwon and Bakhoum).

1.3 Impact of the TME on Anti-Cancer Therapies

Several types of treatment have been proposed and tried in the past years to treat cancer, focusing on directly attacking cancer cells: chemotherapies and targeted therapies. The first one involves the use of drugs designed to kill rapidly dividing cells throughout the body and it is effective in many cases, but it can cause a lot of collateral damage to healthy cells (Anand et al.). Targeted therapies instead are more recent and sophisticated: they interfere with specific molecular pathways that are crucial to cancer cell growth, offering more precise and less toxic options for patients (Min and Lee). Tumors could develop resistance to these types of treatments, by rewiring pathways (i.e. metabolism) or changing behavior, structure, and function of the targeted molecules.

The idea of deploying the immune system as a tool to treat neoplastic disease originated in the nineteenth century (Waldman et al.) and the advent of knockout mouse models provided the necessary technology to experimentally demonstrate a link between immunodeficiency and cancer (Chow et al.). Unlike the previously described treatments that target cancer cells, immunotherapy works by also harnessing and enhancing the patient's own immune system to fight cancer. Hence immunotherapy is based on the concept that the immune system, when properly activated, can recognize and destroy cancer cells. There are many types of immunotherapies, cell- or virus-based (chimeric antigen receptors T-cell therapy or oncolytic virus therapy) and antibody-based (immune checkpoint inhibitors or monoclonal antibodies). Immune checkpoints involve receptors expressed by immune cells that enable dynamic regulation of immune homeostasis. Among this category, the most important and used to treat BC or other types of solid tumors is anti-Programmed Cell Death 1 (PD-1) (McKeage and Perry; Johnson et al.). Interaction between PD-L1 and PD-1, on tumor cells and tumor-infiltrating myeloid cells respectively, results in T cell exhaustion, rendering the T cells inactive (Park and Nam). Checkpoint inhibitors drugs block PD-1 and PD-L1 proteins both on cancer and immune cells, preventing the inactivation of T-cells. They had success in treating melanoma and lung cancer (Johnson et al.). The most recent guidelines to treat a patient with immunotherapy are high expression of PD-L1, high microsatellite instability, and high tumor mutational burden (TMB) (Greco et al.) (Figure 5). These parameters have been selected because they cause an increase in neoantigens production from cancer cells, which in turn can generate an adaptive immune response.

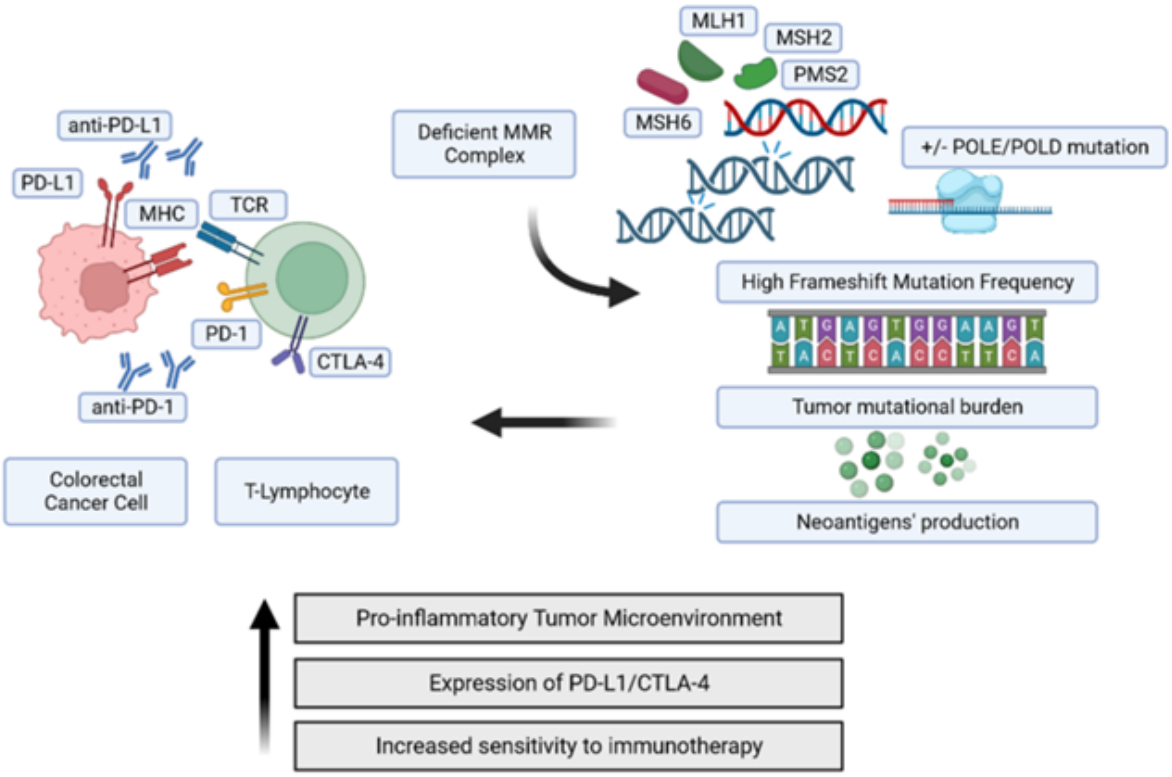

Figure 5: *Current guidelines for treatment with immunotherapies*. This figure shows the effects of a deficient MMR (DNA mismatch repair) system in this example in colorectal cancer cells, which increases the mutation rate, TMB (tumor mutational burden), and neoantigens production, making more susceptible to treatment (Greco et al.).

The major advantage of using immunotherapy is that it may continue to work even after the treatment has ended because the immune system have memory of cancer cells. They are often combined with other treatments like chemotherapy, radiotherapy, or targeted therapies. The combination of chemotherapy, radiotherapy, and targeted therapies with immunotherapies could enhance efficacy but also increase the side effects. Another issue is that some types of cancer with poor presence of active immune system, called "cold" in contrast to the "hot" ones which present a responsive immune system, are resistant to immunotherapies (Jia et al.) (Figure 6). In these cases, the scientific community is trying to find a way to sensitize them to immunotherapy, i.e. recruit immune cells into the tumoral mass, and, in combination with immune checkpoint blockade (ICB), reactivate immune cells to eliminate the tumor.

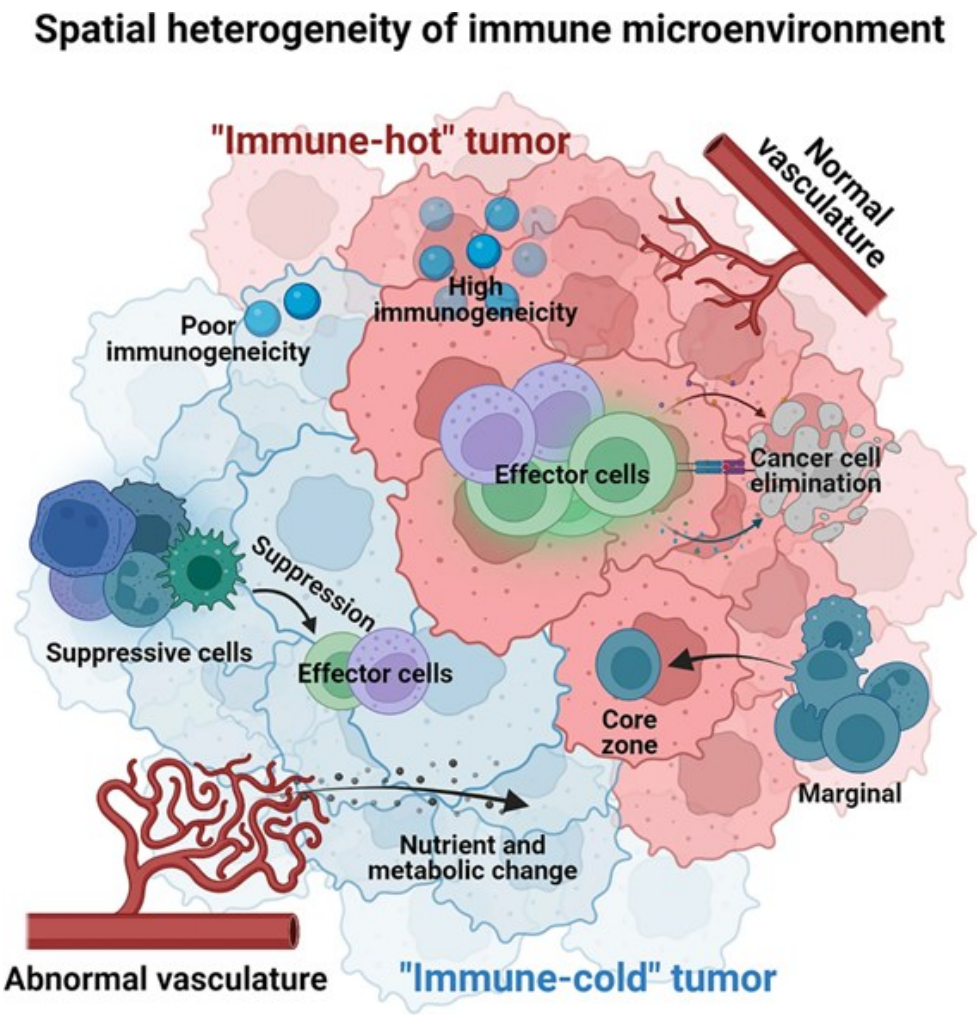

Figure 6: *Spatial heterogeneity of the tumor microenvironment.* TME can be classified as hot or cold depending on the quantity of immune anti-tumor infiltration, which includes mainly CD4+, CD8+ and NK cells (Jia et al.).

1.4 Impact of Oncogenic Signaling on the Evolution of the TME

Significantly, besides affecting cancer cell-intrinsic processes, mutations or deregulation in oncogenes and tumor suppressor genes also influence the composition and activities of the TME, distorting it and favoring pro-cancer functions (Capaci et al.; Gál et al.; Anderson and Simon; Soysal et al.; Mehraj et al.; Mittal et al.; Roulot et al.; Giorello et al.; Yuan et al.).

Among the most important oncogenes, studied also by the hosting lab, there are mutant TP53 (Tumor Protein 53) and YAP1 (Yes1 Associated Transcriptional Regulator). TP53 is the top mutated gene in human cancers and the p53 protein acts as a master TF (transcription factor) sensing stress stimulus, nutrient alterations, activating apoptosis, regulating genome integrity and also cell-cell signaling. It is noteworthy that more than 75% of TP53 mutations are missense point

mutations and some of the resulting mutant p53 proteins are activated and stabilized by different stimuli, and in particular mechanical cues generated by stromal stiffness. Mut-p53 not only loses its normal tumor-suppressive functions, including the inability to activate canonical p53 target genes, but it also exhibits dominant negative effects on the wild-type form of the protein and it acquires new GOF (gain-of-function) oncogenic activities, independent of the wild-type p53, thereby promoting cancer progression (Tombari et al.; Boutelle and Attardi; Kandoth et al.; Berke et al.; Vijayakumaran et al.; Zhu et al.; Mantovani et al.; Alvarado-Ortiz et al.; Ingallina et al.) (Figure 7).

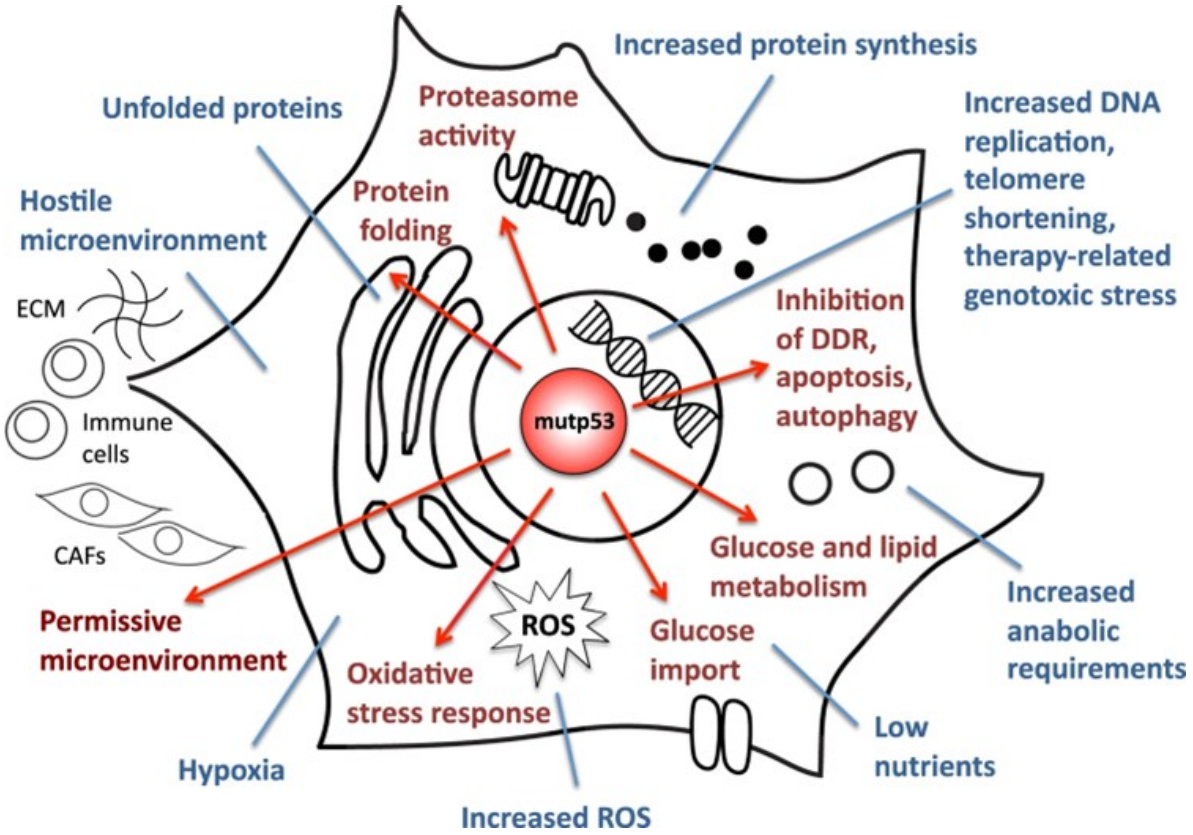

Figure 7: *Schema of mutTP53 functions within the tumor cell.* MutTP53 is involved in several pathways during the evolution of the cancer cell: metabolism, proliferation and inhibition of apoptosis among others (Mantovani et al.).

YAP1 and its interactor TAZ (also known as WWTR1, WW Domain Containing Transcription Regulator 1) are highly related transcriptional regulators pervasively activated in human malignancies and their activation induces increased proliferation, chemoresistance, and metastasis. They are sensors of mechanical and structural features of the cell microenvironment, and upon

activation they shuttle from the cytoplasm in the nucleus, where they recognize the TEAD (TEA domain family members) Transcription Factors (TFs) (Li et al.) (Figure 8).

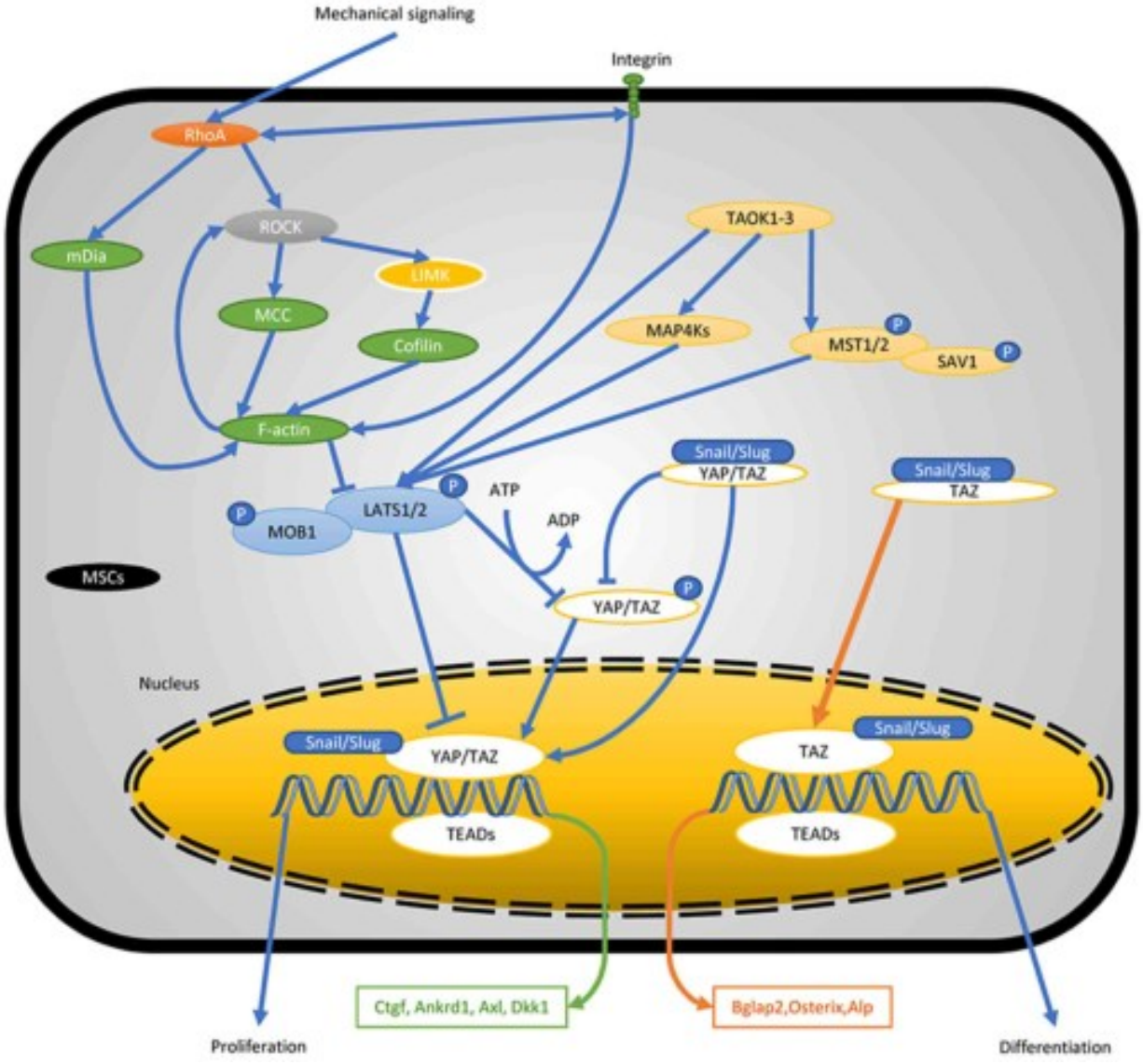

Figure 8: *YAP/TAZ behavior in response to mechanical signaling.* YAP/TAZ is a master regulator of mechanotransduction whose functions rely on translocation from the cytoplasm to the nucleus in response to diverse physical cues (Li et al.).

1.5 The Prolyl-Isomerase PIN1: a Central Hub in Oncogenic Signaling and Therapy Resistance

Peptidylprolyl Cis/Trans Isomerase, NIMA-Interacting 1 (PIN1) is the only known peptidyl-prolyl cis–trans isomerase (PPIase) that specifically recognizes and isomerizes the phosphorylated

Serine/Threonine-Proline (pSer/Thr-Pro) motif. It is involved in many cellular processes, the aberrance of which can lead to both degenerative and neoplastic diseases. Notably, in cancer PIN1 regulates both tumor promoting processes and immune evasion (Chen et al.; Chuang et al.) (Figure 9). Moreover, in a recent study performed in our group in Drosophila melanogaster models, we found that the PIN1 ortholog dodo maintains nuclear type-B Lamin (LamB) structure and anchoring function for heterochromatin protein 1a (HP1a, homolog of HP1α in mammals). By this mechanisms, PIN1/Dodo is crucial to maintain LamB structure, heterochromatin condensation, and thus repression of TEs under mechanical stress that impinge on the nuclear envelope (NE) (Napoletano et al.). Given this relation to mechanical stress and its involvement in immune pathways deregulation in cancer cells, we thought that PIN1 could impact therapy resistance, and its inhibition could be exploited to sensitize resistant tumors to immunotherapies.

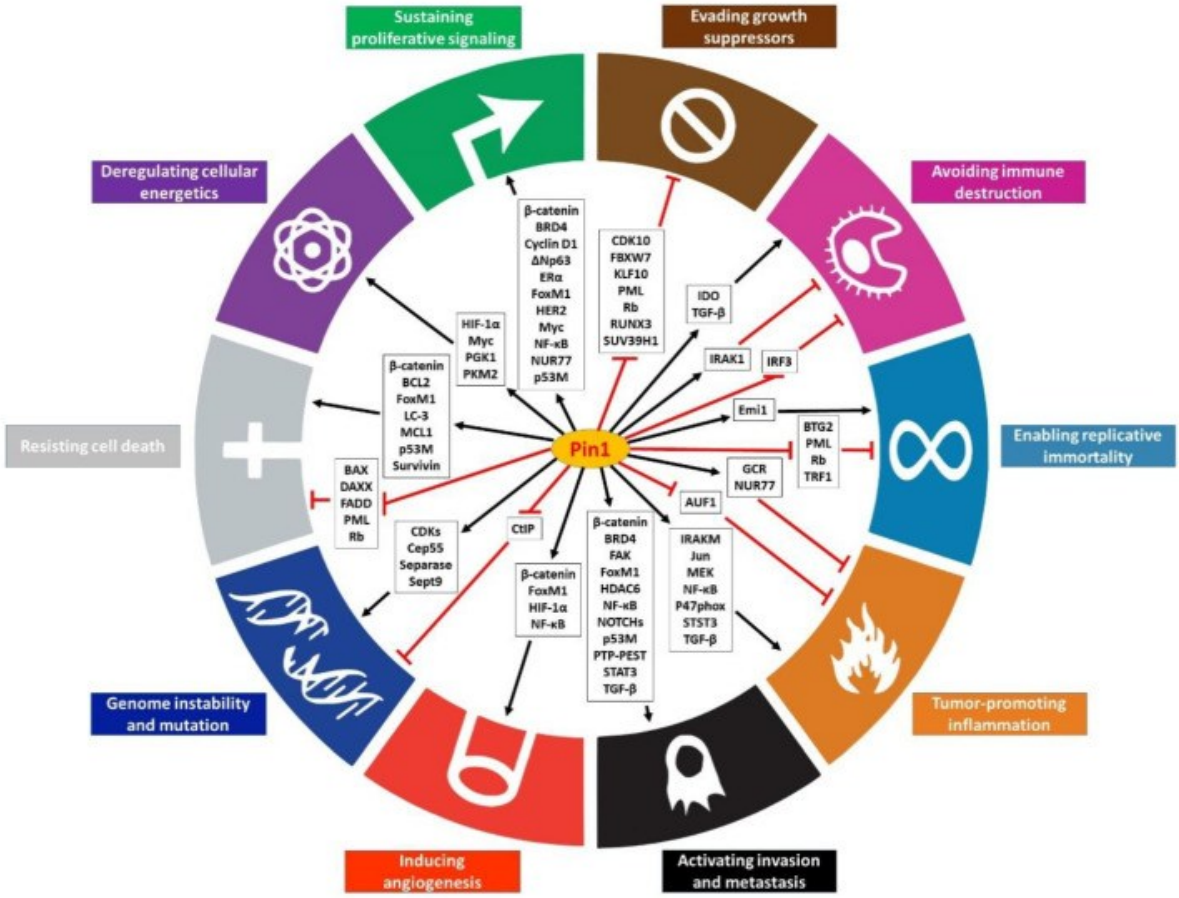

Figure 9: *PIN1 as a central hub for cancer hallmarks.* This figure shows the ten cancer hallmarks and how PIN1 could impact every one of them acting via different pathways (Chuang et al.).

1.6 Bioinformatics Approaches for Precision Medicine

In the rapidly evolving field of multi-omics research, the utilization of multiple methodologies in an integrative manner is of paramount importance for unlocking the profound intricacies of oncological phenomena and propelling progress toward the paradigm of precision medicine (Singer et al.; Akhoon).

Before the advent of bioinformatics approaches, BCs had been classified in subtypes according to immunohistochemical (IHC) staining of Estrogen Receptor (ER), Progesterone Receptor (PR), and Human Epidermal Growth factor receptor 2 (HER2), performed on fresh or frozen BC tissue removed during biopsy. They were divided into ER +/ ER -, PR +/PR -, and HER2 +/HER2 -. In the specific case in which all three of them are absent, the tumor is classified as Triple Negative BC (TNBC).

Throughout the past decades, an extraordinary wealth of diverse omics data has been amassed, spanning multiple domains including genomics, transcriptomics, proteomics, microbiomics, lipidomics, and metabolomics (Lander et al.). This explosion of data has been driven by technological advancements in high-throughput sequencing and other analytical techniques, and in particular:

1. Genomics, which focuses on the study of an organism’s complete set of DNA, including variations that may contribute to cancer development and progression.
2. Transcriptomics, which examines all the RNA transcripts produced from the genome and provides cell gene expression profiles in cancer.
3. Proteomics, which investigates the entire set of proteins expressed by a cell, tissue, or organism.
4. Microbiomics. which studies microorganisms present in a microenvironment, for example in the tumoral one.
5. Lipidomics, which investigates the lipids' presence in biological systems, that can play a crucial role in signaling and metabolism.
6. Metabolomics, which examines the small-molecule metabolites, giving hints regarding altered metabolic pathways in cells.

These comprehensive datasets contain a treasure trove of information that, when integrated, can yield a more comprehensive understanding of cancer biology (Koutsogiannouli et al.). By

analyzing data from various omics disciplines, researchers gain the capacity to unveil intricate relationships, discern hidden patterns, and uncover previously unknown connections within the complex landscape of cancer biology (de Anda-Jáuregui and Hernández-Lemus) (Figure 10).

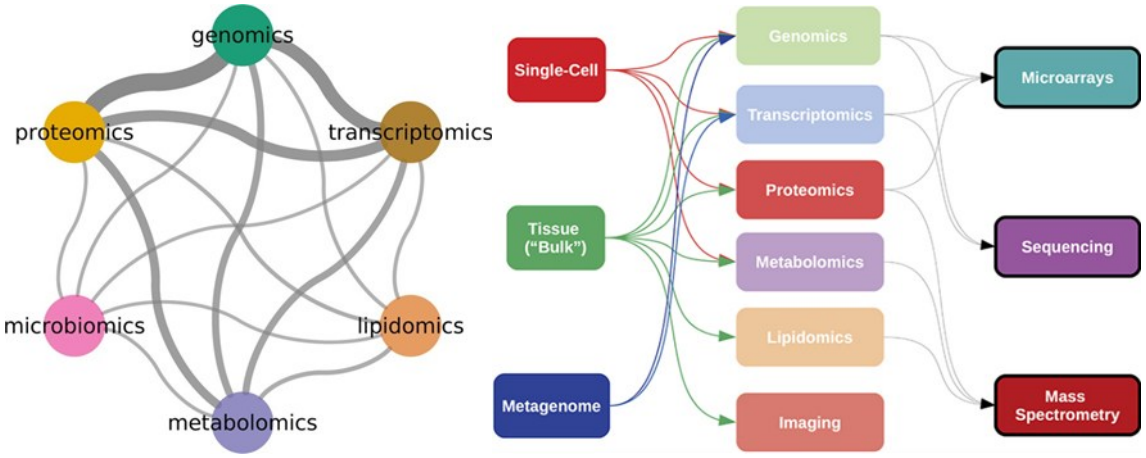

Figure 10: *Omics technology are all interconnected.* The omics techniques have been used together in many papers: the width of the lines in the left graph represents the number of co-occurrences in literature. For each combination of sample type, omic measurement and analytical technology a dedicated pipeline is needed to analyze the data (de Anda-Jáuregui and Hernández-Lemus).

After the generation of this big amount of data, in particular from the now obsolete Microarray technology, scientists developed the PAM50 (Prediction Analysis of Microarray 50) Gene expression classifier, a list of 50 genes measured parallel with 8 housekeeping genes (Prat et al.; Dank et al.). It classifies BCs into Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like subtypes. Usually, the Basal-like tumors are comparable to the TNBC IHC subtype, whereas Luminal A and B are equally distributed among ER+/PR+/HER2- and ER+/PR+/HER2+. Finally, the HER2-enriched usually have an ER-/PR+/HER2+ genotype. The conventional PAM50 subtyping procedure has been widely recognized as a crucial aspect of BC research, providing valuable intrinsic subtypes to the scientific community. However, it is common for the PAM50 subtypes to fall short of fully capturing the variations and heterogeneity present within different BC subtypes, sometimes resulting in misclassification (Raj-Kumar et al.).

Within this multi-omics framework, the advent of innovative Next generation sequencing (NGS) (Satam et al.) tools such as RNAseq, ChIPseq (Chromatin Immunoprecipitation Sequencing), single-cell RNA sequencing (scRNAseq), and spatial transcriptomics technology have revolutionized our ability to study cancer and its microenvironment. RNAseq and scRNAseq serve

as indispensable approaches for exploring the transcriptional profiles of biological samples, each offering unique advantages in unraveling gene expression in cancer:

1) Bulk RNA sequencing allows us to quantify gene expression levels, identify novel transcripts and splice variants, detect gene fusions, analyze differential gene expression between conditions, and explore non-coding RNAs, such as long non-coding or microRNAs. It is an evolution of Microarray technology, which could not identify non-coding RNAs due to the limitation of the array platforms (Rao et al.). The major limitation of RNAseq is that it is unable to resolve cell-type specific expression patterns in heterogeneous samples.
2) scRNAseq allows dissecting the molecular characteristics of individual cells, offering unprecedented resolution to unravel cellular heterogeneity. This technique permits the identification and characterization of cell populations within a heterogeneous sample, the reconstruction of developmental trajectories and lineage relationships, and the discovery of new cell types or states(Jovic et al.; Heumos et al.).

This increase in resolution is particularly crucial in cancer research because tumors can contain several cell types (immune, stromal, and vasculature cells) and tumor cells at various stages of progression. With the scRNAseq technology, we can identify Genes Differentially Expressed (DEGs) and enable Gene Set Enrichment Analysis (GSEA) within each identified cell population.

Additionally, scRNAseq offers insights into direct or indirect communications between different cell populations based on the phenotypic condition of the samples, for instance by analyzing expression of ligands and cognate receptors. This aspect is particularly powerful for understanding the complex interactions within the TME, for instance, we can detect de-regulated ligand-receptor interactions.

ChIPseq provides insights into protein-DNA interactions and epigenetic regulation and can be integrated with RNAseq and scRNAseq. This technology allows us to map the genomic locations of transcription factors (TFs) binding sites, identify regions of active transcription or repression, and understand how changes in the epigenetic landscape contribute to aberrant gene expression in cancer.

The newest addition to this toolkit, spatial transcriptomics, adds another layer by preserving spatial information. With this technique, we can visualize and quantify gene expression patterns, identify

spatially restricted gene expression programs, understand how the physical location of cells influences their transcriptional state, and map the distribution of different cell types within the TME.

The integration of these remarkable tools, including RNAseq, ChIPseq, single-cell RNAseq, and spatial transcriptomics, has opened up unparalleled avenues for the comprehensive exploration of cancer biology. By harnessing data from multiple omics disciplines, researchers can unlock novel molecular signatures, identify potential biomarkers, and pave the way for the development of personalized therapeutic strategies. Such integrative approaches are indispensable for harnessing the full potential of multi-omics data, advancing our knowledge, and driving progress toward improved cancer diagnosis, treatment, and enhanced patient outcomes. Moreover, in recent years, a particular focus has been put on the study of expression of genome sequences other than protein-coding genes and TEs concerning cancer development, and many tools have been developed to study this dark part of the genome. This massive fraction of the genome could be relevant for comprehending a new shade of the TME and tumor cell complexity.

2.AIM OF THE THESIS

Based on these premises, the central hypothesis of this Ph.D. project is that there exists key determinants able to sense and cooperate with the TME and in particular the ECM stiffness in order to confer tumors resistance to therapies. To assess this hypothesis, BCs classification could be improved by classifying patients according to molecular signature scores. Based on this hypothesis, we divided the project into the following main tasks:

- Find signatures/axes that classify patients and identify key determinants sensitive to mechanosignaling:
    - Retrieving molecular and phenotypic signatures from transcriptomic analysis (RNAseq) in experimental models.
    - Analyzing the retrieved signatures in bulk RNAseq from patients.
- Investigate the pathways modulated by these determinants in bulk RNAseq and scRNAseq data.
- Search for potential immunogenic neoantigens.

Among determinants of therapy resistance, we focused on the prolyl-isomerase PIN1, which was shown to be overexpressed in aggressive BC and to support resistance to chemotherapy (Rustighi et al.). We aimed to establish possible correlation between PIN1 activity and both fibrotic and immune cold BCs, assess the impact of PIN1 in anti-tumor immune response, and whether PIN1 activity could be used to refine BC classification in order to better predict response to therapy.

3.RESULTS

We proceeded to exploit transcriptional signatures that could be used to analyze the correlation between PIN1 protein expression, ECM stiffness, and immune cell status. To generate the signatures, we proceeded through three steps:

1. Download reference sequences of genomes and annotation of protein-coding and non-coding sequences of Human (Homo sapiens), Mouse (Mus musculus) and Drosophila melanogaster species.
2. Set up an automatic pipeline for alignment of sequenced RNA reads with reference genomes
3. Generate molecular and phenotypic signature scores from RNAseq experiments (Figure 11).

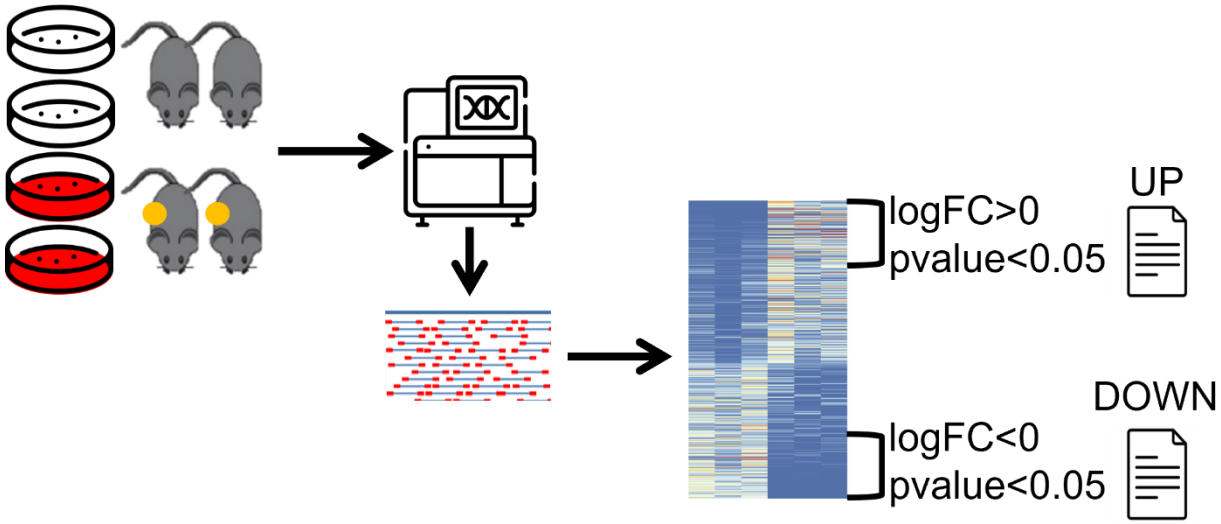

Figure 11: *pipeline to retrieve signatures from sequencing data.* To retrieve molecular and phenotypic signatures, we took as input the raw fastq data from experiments, aligned them and calculated the DEGs, and took the ones with pvalue < 0.05 and log2FC>0 for the upregulated genes, and log2FC<0 for the downregulated ones.

3.1 Download Genome Sequences, Gene and TEs Annotations

We began by acquiring a comprehensive set of genomic data. We downloaded all the genomes sequences, genes and TEs annotations from the online databases ENSEMBL (Cunningham et al.) and UCSC (Kent et al.). These databases are gold standards for their curated, high-quality genomic information.

Since we wanted to explore the often-overlooked regions of the genome containing TEs, we conducted an extensive literature review to identify state-of-the-art pipelines designed for detecting these genomic regions in RNAseq data and we have found three pipelines specific on TEs:

- *RediscoverTE* (Kong et al.): This software was initially designed only to detect TEs in human samples and never adapted to other organisms. Recognizing the potential of this tool, we took the initiative to extend the pipeline to accommodate mouse data.
- *SQuIRE* (Yang et al.): This pipeline has been previously used in a publication conducted in the hosting lab in Drosophila melanogaster (Napoletano et al.). To ensure its applicability, we validated the results reported in the paper using the original Drosophila data.
- *Telescope* (Bendall et al.): as the most recently developed pipeline, Telescope has gathered attention for its advanced capabilities and its superiority has been corroborated by several statical analyses (Lanciano and Cristofari).

All three selected pipelines require FASTQ or BAM files as input, necessitating access to raw sequencing data rather than processed results. To this end, we initiated a formal request to the The Cancer Genome Atlas (TCGA) consortium to gain access to this type of data and then we applied REdiscoverTE to quantify the amount of TEs present among the Breast invasive carcinoma (BRCA) patients.

TCGA comprises complete data regarding 921 patients, including clinical information, copy number alterations, miRNA profiles and RNA sequencing data. We compiled Table 1 summarizing key clinical information about the BC patients included in this study.

Clinical data	Description	Number of patients
ethnicity	An individual's self-described social and cultural grouping.	Not hispanic or latino: 745 NR: 147 Hispanic or latino: 29
gender	Gender is described as the assemblage of properties that distinguish people on the basis of their societal roles.	Female: 910 Male: 11
race	It is characterized by shared heredity, physical attributes and behavior, and in the case of humans, by common history, nationality, or geographic distribution.	White: 632 Black or african American: 151 NR: 79 Asian: 58 American indian or alaska native: 1
ajcc_pathologic_m	Code to represent the defined absence or presence of distant spread or metastases.	M0: 769 MX: 128 M1: 18 cM0 (i+): 6
ajcc_pathologic_n	The codes that represent the stage of cancer based on the nodes present.	N0: 278 N0 (i-): 133 N0 (i+): 25 N0 (mol+): 1 N1: 102
ajcc_pathologic_stage	The extent of cancer, especially whether the disease has spread from the original site to other parts of the body.	Stage I: 77 Stage IA: 72 Stage IB: 5 Stage II: 6

		Stage IIA: 317 Stage IIB: 208 Stage III: 1 Stage IIIA: 127 Stage IIIB: 23 Stage IIIC: 47 Stage IV: 17 Stage X: 11 NR: 10
ajcc_pathologic_t	Code of pathological primary tumor to define the size or contiguous extension of the primary tumor.	T1: 36 T1a: 1 T1b: 13 T1c: 189 T2: 543 T2a: 1 T2b: 1 T3: 97 T4: 8 T4b: 26 T4d: 3 TX: 3

Table 1: *Clinical TCGA BRCA data. For each tumor clinical data, we have reported the number of cases belonging to each category. When NR is present the data is Not Reported.*

3.2 Set Up an Automatic Alignment Pipeline

To automatically analyze the RNAseq data for which we aimed to retrieve transcriptional signatures, we developed a comprehensive pipeline, integrating several state-of-the-art bioinformatics tools. The pipeline is designed to process raw FASTQ data and produce a multifaceted analysis including differential gene expression, pathway enrichment and TEs abundance quantification. Here is a breakdown of the pipeline components:

1) TRIMGALORE (Martin): this is a wrapper tool around Cutadapt and FastQC. It removes low-quality base calls, trims adapter sequences and removes very short reads that could lead to incorrect alignment.
2) STAR (Spliced Transcripts Alignment to a Reference) (Dobin et al.): we selected STAR for read alignment because of its ability to handle splice junctions efficiently. We performed both the alignment of coding and non-coding regions, and specifically in the second case we have used specific parameters for capturing multimapping reads, since the repetitive nature of TEs:
    a. --winAnchorMultimapNmax 100: Allows up to 100 mappings per read within the alignment window.
    b. --outFilterMultimapNmax 100: Reports up to 100 alignments for multimapping reads.
    c. --alignIntronMax 100000: Sets the maximum intron length to 100,000 bases, accommodating most known introns.
    d. --outFilterMismatchNoverLmax 0.04: Allows for a mismatch rate of up to 4%, balancing sensitivity and specificity.
3) DESeq2(Love et al.): for differential expression analysis we have selected DESeq2, a robust and widely used R package. It normalizes count data to account for in-between-samples sequencing depth differences, estimates dispersion and fits a negative binomial model for each gene, performs statistical tests to identify DEGs and calculate the false discovery rate.
4) GSEA (Mootha et al.; Subramanian et al.): it takes the list of genes ordered by their differential expression statistics from DESeq2, evaluates the enrichment of predefined gene sets and calculates enrichment scores.

5) Telescope (Bendall et al.): the final component of the pipeline is Telescope, designed for quantifying TE expression from RNAseq data. It takes the BAM files produced by STAR as input, employs a probabilistic model to assign ambiguously mapping reads to their most likely source TE loci and provides locus-specific abundance estimates for each TEs.

The integration of these tools into a single, automated pipeline, shown in Figure 12, ensures that all samples are treated identically, reducing technical variability and provides a holistic view of transcriptional changes, from individual genes to pathway and repetitive elements.

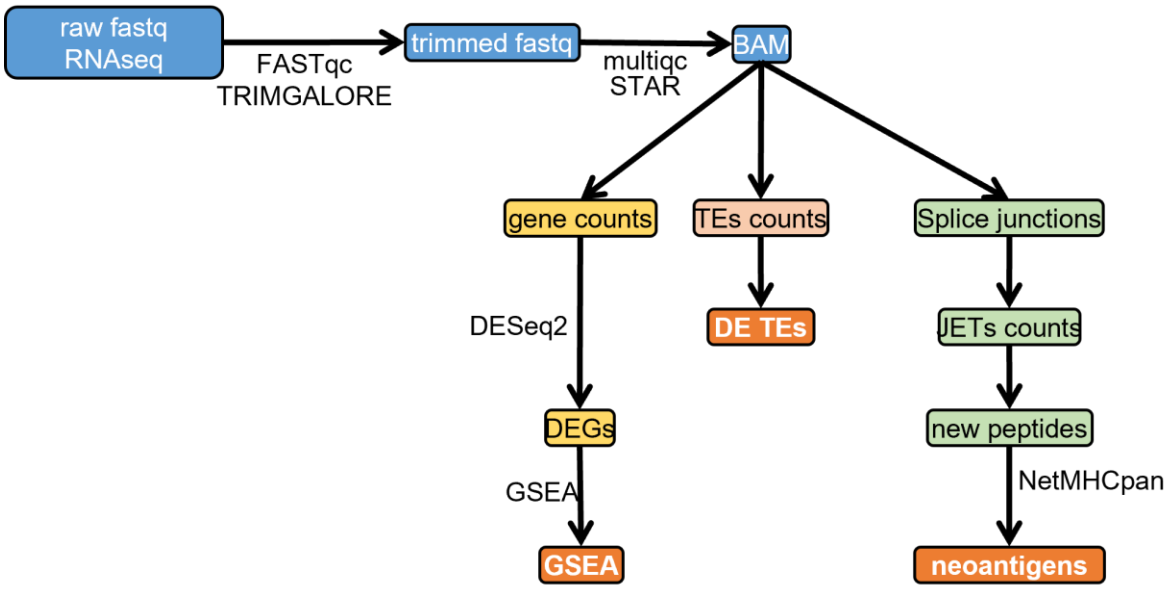

Figure 12: *pipeline for the analysis of raw FASTQ data.* Beginning from the raw sequencing data, we have reported all the tools used and intermediate results obtained in my pipeline.

3.3 Generation of Molecular and Phenotypic Signatures Scores from RNAseq Experiments

Signatures of different immune cell populations and of main mechanotransducers (i.e. YAP) were already published and available.

In particular, we have downloaded the signature of CONVENTIONAL DC (López et al.), the one regarding MHC-II, T cells and B cells (Bagaev et al.) and the TILs one (Ballot et al.). The YAP targets genes activity signature has been retrieved from (Zanconato et al.).

We generated signatures of ECM stiffness and PIN1 activity by applying the aforementioned pipeline on RNAseq data deriving from cells grown at different ECM stiffness and cells with PIN1 ablation, respectively. Specifically, the upregulated and downregulated DEGs with a pvalue less than 0.05 were the chosen gene lists.

Since clinically the tumor tissue density is measured through mammographic data, and we did not have direct access to them from the online datasets that we were analyzing, we chose to retrieve a signature from the data of (Watson et al.) as a proxy of the stiffness phenotype. To generate this signature, TNBC tissue cell cultures expressing mutTP53 SUM15 have been grown in different collagen densities (from 0.5 kPa to 8.0 kPa) and ensuing elastic moduli.

To generate the signature for PIN1 activity, we have re-analyzed two RNASeq experiments conducted on MDA-MB-231 cell line in which authors silenced PIN1 with the KPT-6566 ((Campaner et al.) and (Girardini et al.)). Moreover, four RNAseq experiments (two made in PIN1KO 4T1 cell line, prepared with ribozero and polyA sequencing library respectively, one in a mouse model harboring MMTV-PyMT BC cells with tamoxifene-induced conditional PIN1KO and one in MCF7 BC cell line treated with two different siPIN1, named #1 and #7) have been set up in the laboratory.

Signature ID	Signature name	Source	Overview	Number of genes
1	conventional DC activation	(López et al.)	activation of Dendritic Cells	5
2	MHC-II	(Bagaev et al.)	regulate presentation of cancer-associated peptides to the tumor-reactive CD4+ T cells	9
3	T cells	(Bagaev et al.)	includes surface markers, components of TCR, transcription factors and others.	11
4	B cells	(Bagaev et al.)	comprises genes characteristic for B cells, including surface markers, transcription factor and others.	13
5	TILs	(Ballot et al.)	Tumor infiltrating lymphocytes	10
6	YAP-TAZ-TEAD direct target genes from Zanconato et al.	GSE66082	MDA-MB-231 treated with siYAP/TAZ or siCtrl	381
7	stiffness-related signature from Watson et al.	GSE127887	SUM159 plated on soft (0.5 kPa) and stiff (8.0 kPa) condition	5703 UP/ 5480 DOWN

8	PIN1 activity from Campaner et al.	GSE84909	MDA-MB-231 transfected with PIN1 siRNA or control siRNA	1721 UP/ 193 DOWN
9	PIN1 activity from Girardini et al.	GSE26262	MDA-MB-231 transfected with PIN1 siRNA or control siRNA	3349 UP/ 3724 DOWN
10	PIN1 activity from KO	in lab made experiment	4T1 murine mammary carcinoma cell line edited with PIN1KO and processed with ribozero library	242 UP/ 456 DOWN
11	PIN1 activity from KO	in lab made experiment	4T1 murine mammary carcinoma cell line edited with PIN1KO and processed with polyA library	367 UP/ 420 DOWN
12	PIN1 activity from Tamoxifene-induced PIN1KO	in lab made experiment	PyMT murine mammary carcinoma cell line edited with PIN1 tamoxifene-inducible KO	63 UP/ 181 DOWN
13	PIN1 activity from siPIN1#1	in lab made experiment	MCF7 BC cell line	234 UP/ 664 DOWN

14	PIN1 activity from siPIN1#7	in lab made experiment	MCF7 BC cell line	282 UP/ 443 DOWN

Table 2: *Molecular or phenotypic signatures. For the one coming from GEO datasets, we specified also the GSE identifier.*

In the field of transcriptomics and gene signature analysis, there is ongoing debate regarding the optimal method for assigning a signature score to samples or patients. We undertook a comprehensive approach to evaluate and compare different scoring methods. To address this methodological question, we decided to implement and compare four approaches for signatures score calculation:

- Method A: sum of concordant Z-score-transformed signature genes expression. It focuses only on the genes whose expression changes are in the same direction as the original signature. This approach emphasizes the contribution of genes that behave consistently with the signature, reducing noise from genes with inconsistent behavior (Figure 13A).
- Method B: difference of sums of concordant and discordant Z-score-transformed signature genes expression. It considers both upregulated and downregulated genes, subtracting the contribution of discordant genes from concordant ones. By incorporating both concordant and discordant genes, this method provides a more balanced view of the overall signature activity (Figure 13B).
- Method C: difference of means of concordant and discordant Z-score-transformed signature genes expression. It is like method B, but it normalizes by the number of genes in each category. This approach aims to balance the contribution of up- and down-regulated genes, preventing bias due to unequal numbers of genes in each category (Figure 13C).
- Method D: singscore R package score (Foroutan et al.). This method utilizes the singscore R package, which implements a rank-based scoring system. It offers a non-parametric, rank-based approach that may be more robust to outliers and differences in gene expression measurement scales across datasets (Figure 13D).

To compare these four methods, we have chosen to calculate the correlations between the PIN1 mRNA levels and the PIN1 activity signature score from the list in Table 1 (union between signatures 13 and 14) and check which method has the highest correlation and the best one was singscore (Figure 13).

Given these results, we decided to proceed with the singscore method, also due to the method's theoretical advantages. The package offers two key functions that we utilized in my analysis:

1) *rankGenes*: it takes an expression matrix as input and returns a matrix where genes are ranked for each patient. This step normalizes gene expression across samples, making scores more comparable across datasets or experimental conditions.
2) *simpleScore*: this function calculates the actual signature score. It accepts the ranked gene matrix produced by *rankGenes* as input and offers flexibility in signature definition, allowing for both single-direction and bi-directional signatures (Table 2).

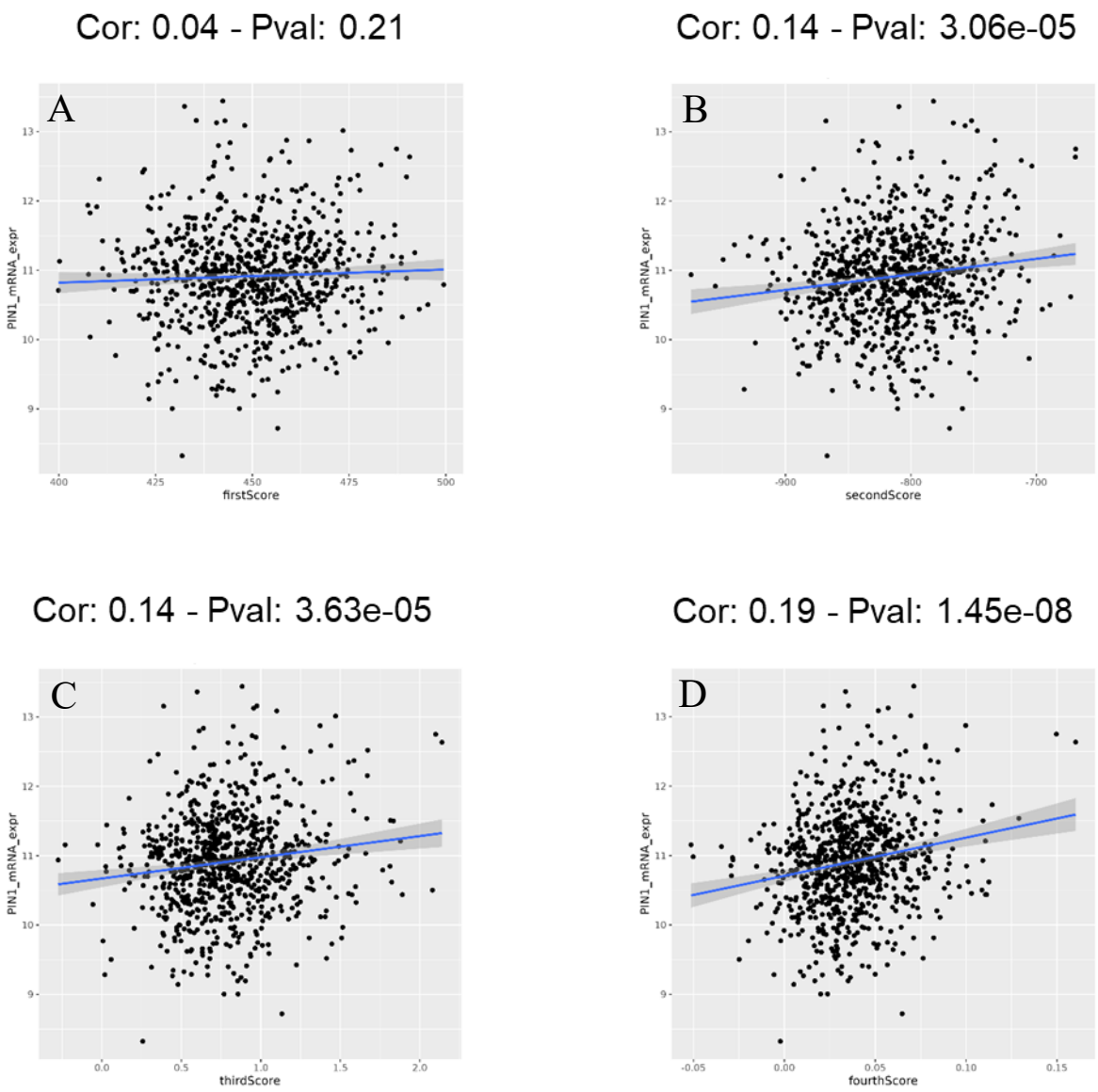

Figure 13: *comparison between score assigning methods.* Correlation plots between PIN1 activity signature score (on the x axis) and PIN1 mRNA expression (on the y axis) in TCGA BC patients. In A the score is calculated using method A, in B using method B, in C using method C and in D using method D. Each point represents a patient, and the blue line represents the regression line with its boundaries.

3.4 Annotation and Signatures Correlation of BC Patients

Using the calculated scores derived from the singscore method, we decided to investigate the intricate biological relationships between PIN1 activity, ECM stiffness and immune cell types in BC patients. This analysis leveraged the rich dataset from TCGA BRCA cohort, allowing for a population-level examination of these complex interactions.

To delve deeper into the relationship between these signatures we employed a correlation analysis using the Pearson correlation coefficient. This statistical measure quantifies the linear correlation between two variables, producing values between +1 and –1, where 1 indicates a perfect positive correlation, 0 indicates no linear correlation and –1 indicates a perfect negative correlation. For this analysis, we utilized the *rcorr* function from the Hmisc (Harrell Jr) (version 5.1-3) R package. This function is useful as it computes a matrix of Pearson correlation coefficients for all possible pairs of columns in a given matrix, along with the corresponding p-values. The correlation analysis revealed several key findings that align with previous biological evidence.

Figure 14A shows a positive correlation between the PIN1 activity (union between signatures 13 and 14) and ECM stiffness signature scores. This finding corroborates existing evidence suggesting a mechanistic link between ECM remodeling and PIN1 in cancer. This important enzyme may react to changes in ECM stiffness, a feature associated with tumor progression and metastasis.

ECM stiffness signature scores showed a negative correlation with the signature scores of anti-tumoral immune cell types (DCs, T cells, B cells, TILs, and MHC-II presenting cells). This observation aligns with the growing evidence that increased ECM stiffness can generate a physical and biochemical barrier that impedes immune cell infiltration.

Finally, a negative correlation was found between PIN1 activity and the anti-tumoral immune cell types cited above. This inverse relationship suggests that higher PIN1 activity in BC may be associated with a suppressed anti-tumor immune response. The negative correlation with MHC-II presenting cells is particularly intriguing, as it may indicate a mechanism by which high PIN1 activity could contribute to immune evasion through reduced antigen presentation.

The concordant negative correlations of both PIN1 and ECM stiffness with anti-tumoral immune signatures indicate that these factors may work in concert to generate an immunosuppressive microenvironment. This could occur through direct effects of alterations of the physical proprieties

of the tumor stroma. This, moreover, suggests that targeting PIN1 or ECM proprieties could have multifaceted effects, potentially modulating the structural aspects tumor and its immune microenvironment. However, it is important to note that correlation does not imply causation, and the experiments in Figure 14B and 14C, performed on MCF10 DCIS.COM and MCF7 cell line respectively, confirmed that increasing stiffness increases the nuclear localization of both YAP and PIN1, and in soft condition, upon mechanical stress, PIN1 becomes mainly nuclear.

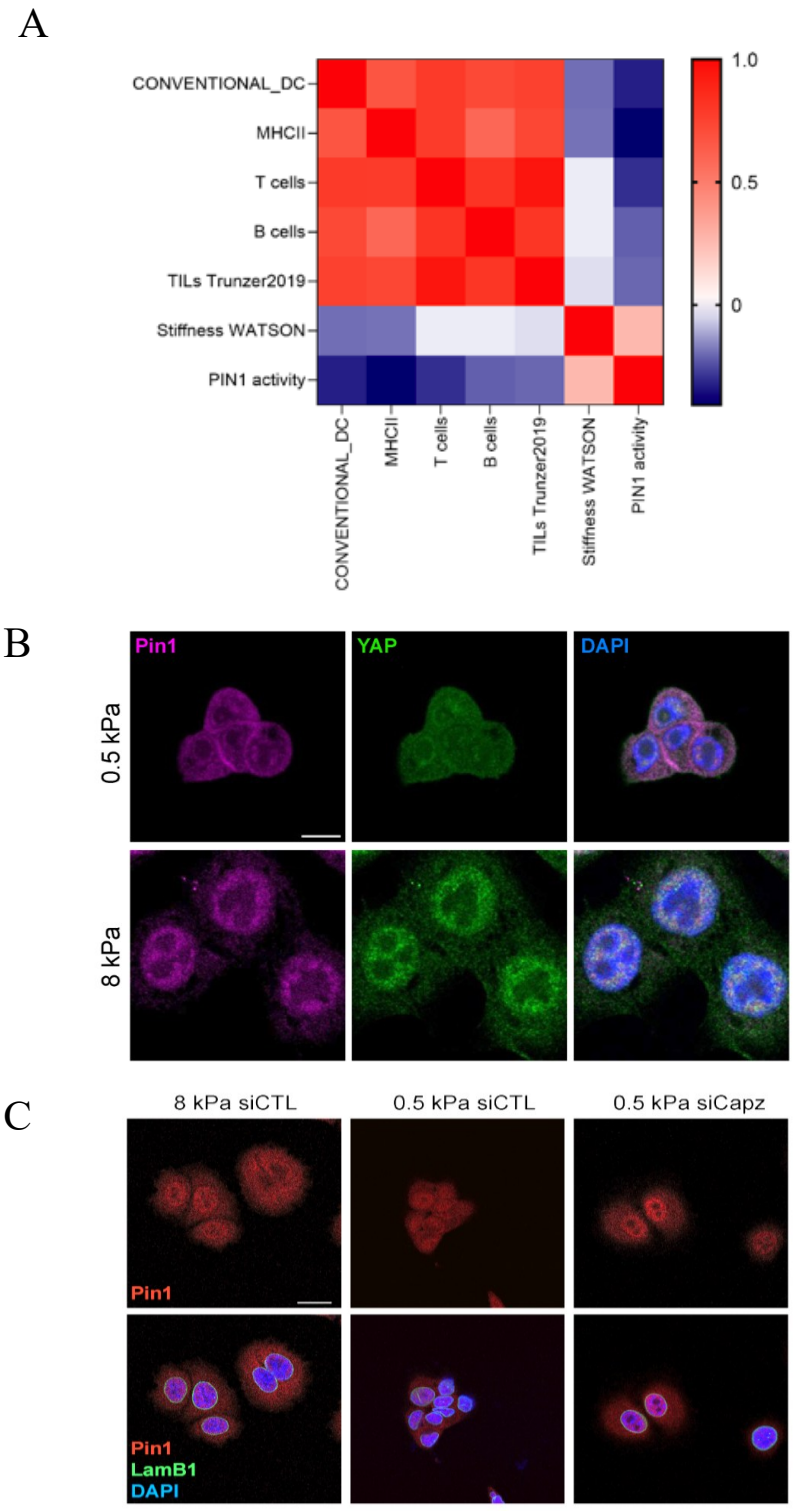

Figure 14: *PIN1 correlates with stiffness and anti-correlates with anti-tumor immune cells presence.* A) Correlation heatmap among 5 anti-tumor immune cells, stiffness and PIN1 activity signature scores. B) PIN1 and YAP cellular localization in MCF10 DCIS.COM in soft and stiff conditions. C) PIN1 localization in MCF7 in stiff, soft and soft upon mechanical stress conditions. MCF10DCIS.com was derived from a xenograft lesion following two trocar passages of the premalignant cell line MCF-10AT while MCF7 is a human BC cell line with estrogen, progesterone and glucocorticoid receptors. Courtesy of Rebecca Bertolio.

3.5 PIN1 Knockdown Reactivates an Innate Immune Response and Derepresses Heterochromatin

To elucidate whether PIN1 ablation could be associated with increased immune system activation, we conducted a comprehensive analysis of the transcriptome from the previously described MCF7 cell line treated with PIN1 siRNAs (siPIN1). This approach allows for a direct examination of the cellular changes induced by PIN1 ablation, potentially revealing the molecular basis of correlations observed in patient data. MCF7 cell line had been instrumental in deriving the PIN1 activity signature. The gene expression data underwent rigorous examination, beginning with principal component analysis (PCA), a powerful statistical technique used to reduce the dimensionality of complex datasets while retaining most of the variation. The results of the PCA were visualized in a multidimensional plot (Figure 15A). Although the PCA did not reveal a striking separation between the treated samples and untreated control samples, suggesting some degree of overlap in their gene expression profiles, it provided a foundation for further investigation. To delve deeper into the molecular changes induced by PIN1 ablation, we performed GSEA comparing the control samples to those treated with siPIN1#1. GSEA is a powerful computational method that determines whether a predefined set of genes shows statistically significant concordant differences between two biological states. In this case, those states were the control condition and the PIN1-ablated condition.

The strength of this method is that it can detect coordinated changes in gene expression that might be missed by single-gene analysis methods. It takes as input the ranked gene list of DEGs and their statistic together with a list of gene sets, retrieved from various sources such as the MSigDB, which includes collections like hallmark gene sets, curated gene sets, and Gene Ontology (GO) terms. These gene sets represent various biological pathways, cellular processes, and molecular functions. Next, we ran the GSEA algorithm, which involved several steps:

- Calculating an enrichment score for each gene set, which reflects the degree to which the set is overrepresented at the top or bottom of the ranked list.
- Estimating the statistical significance of the enrichment score by using permutation testing.
- Adjusting for multiple hypothesis testing to calculate the false discovery rate.

The analysis was performed iteratively, with careful attention paid to various parameters such as the choice of ranking metric, the number of permutations, and the threshold for statistical

significance. This iterative process helped ensure the robustness and reliability of our results. The output of the GSEA included:

- Enrichment plots for each significant gene set, visually representing the distribution of genes in the set across the ranked list.
- Normalized enrichment scores (NES) indicating the degree of enrichment.
- False discovery rates (FDR) and nominal p-values to assess the statistical significance of the enrichment.

We selected the HALLMARK, GO and REACTOME databases from MSigDB and intriguingly, the GSEA revealed a significant enrichment of both innate and adaptive immune responses, with a particular emphasis on IFN and antigen-related pathways (Figure 15B).

This finding suggests that PIN1 ablation may indeed trigger a broader immune activation within the cells, potentially altering their interactions with the immune system. The connection between PIN1 ablation and immune activation led us to explore the cGAS/STING1 pathway, a critical component of the innate immune response. This pathway has been increasingly studied in both healthy and tumor cells, where it has been shown to stimulate the activation of interferon signaling. Previous research had indicated that TEs might play a role in activating this cytosolic DNA sensing pathway. This intriguing possibility prompted me to investigate the regions of the genome where these elements reside. TEs potential role in activating innate immune responses in cancer cells represents a novel intersection of genomics and tumor immunology. TEs are DNA sequences capable of moving within the genome. These genomic nomads have coexisted with our genetic material for millions of years, undergoing continuous evolution and diversification. This long-term coevolution has resulted in a remarkable variety of TEs, each with unique characteristics and potential impacts on cellular function. To better understand the landscape of TEs, it's essential to recognize their classification. TEs are primarily divided into two major classes: class 1 and class 2. Class 1 elements, also known as retrotransposons, employ a distinctive "copy-and-paste" mechanism for their propagation. This process involves transcription of the TE into RNA, followed by reverse transcription back into cDNA, which is then inserted into a new location in the genome. This mechanism allows retrotransposons to increase their copy number with each transposition event. Class 1 elements can be further subdivided into long terminal repeat (LTR) retrotransposons, which include endogenous retroviruses, and non-LTR retrotransposons, such as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Each of these subtypes

has unique structural features and potential functional impacts on the host genome. In contrast, class 2 elements, or DNA transposons, utilize a "cut-and-paste" mechanism. These elements can excise themselves from one genomic location and reinsert at another, typically without increasing their copy number. However, if this process occurs during DNA replication, it can lead to an increase in element copy number. The diverse nature of TEs and their potential to influence gene expression and genomic stability make them an intriguing area of study, particularly in the context of immune activation and cancer biology. As we described in the above chapter, the pipeline that we decided to use to investigate the TEs abundance is Telescope and, in selected settings, we found that there is broad deregulation of TEs expression, involving both LINE, SINE, DNA and LTR TEs, with the majority of the upregulated ones involving SINEs (1667 loci) and LINEs (686 loci) families (Figure 15C). This suggests that PIN1 might play a crucial role in suppressing TEs expression under normal conditions. The upregulation of these elements upon PIN1 ablation could contribute immune activation through several mechanisms such as direct activation of cGAS/STING1 during the retrotransposition process, genomic instability and novel proteins or peptides produced from TE transcripts could serve as potential neoantigens, stimulating the adaptive immune response. The predominance of SINE upregulation might reflect a global change in chromatin state or transcriptional regulation following PIN1 ablation.

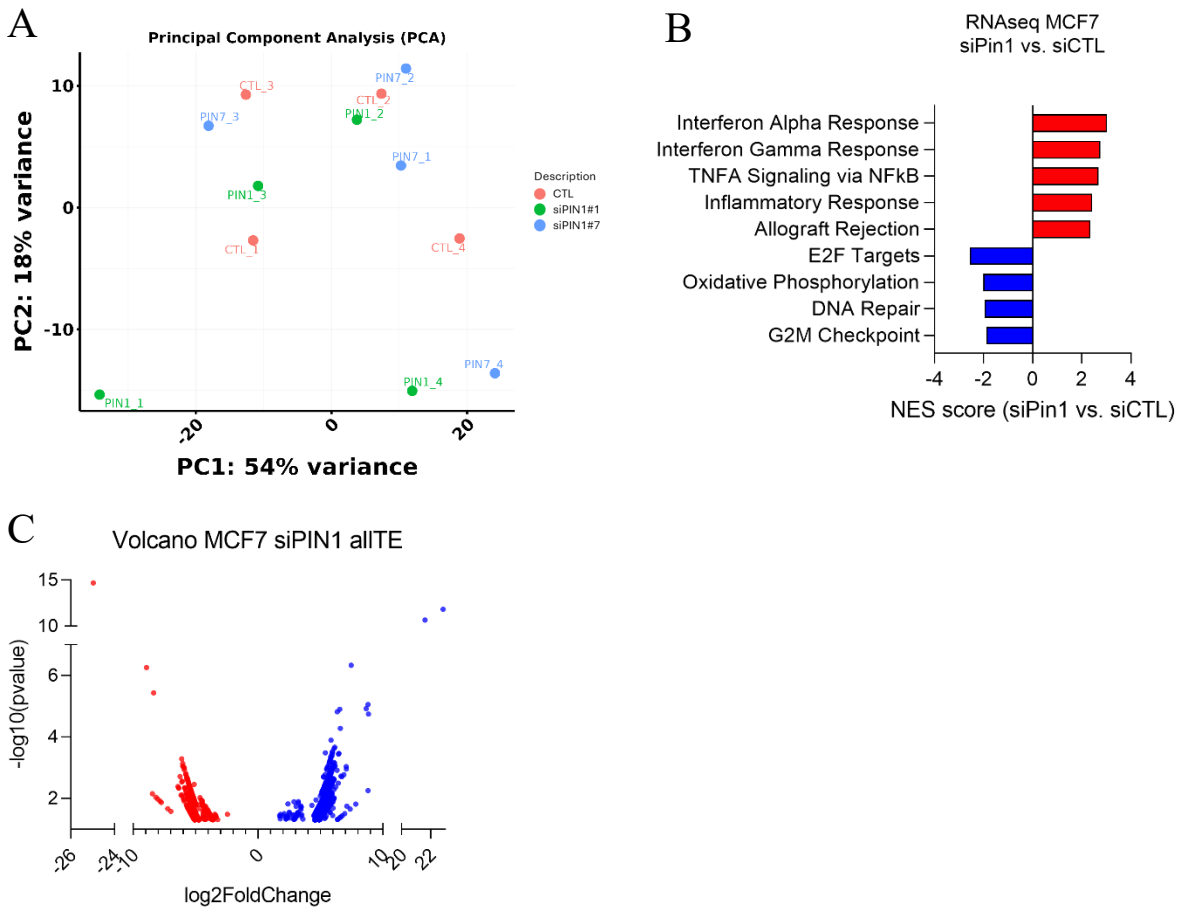

Figure 15: *PIN1 silencing activates an innate immune response and deregulates heterochromatin*. A) PCA of 12 samples derived from the RNAseq analysis of MCF7 BC cell line treated with either siCTL (in red), siPIN1#1 (in green) or siPIN1#7 (in blue). B) PreRanked GSEA of the DEGs calculated using the R package DESEq2: on the y axis there are the categories while on the x axis the NES score in the comparison siPIN1 vs siCTL. C) Volcano plot of the expressed TEs, showing on the y axis the negative log10 of the pvalue and on the x axis the log2 of the FC (Fold Change).

3.6 Ablation of PIN1 in the 4T1 Mouse Model Generates Immunogenic Neoantigens

The innovative approach to investigate the potential neoantigen-generating capacity of TE deregulation following PIN1 ablation represents a cutting-edge intersection of cancer biology, genomics, and immunology. To elucidate whether TE deregulation could produce novel proteins or peptides and, in turn, potential neoantigens upon PIN1 ablation, we applied a recently developed pipeline from the group of (Burbage et al.; Merlotti et al.). This pipeline allows the detection of TE-derived neoantigens generated from non-canonical splice junctions between exons and TEs (JETs), coming from both intra- and inter-exons regions. Reads coming from the RNAseq experiment were aligned to the reference genome (MM39) using STAR, unannotated junctions

were extracted and those that included both a TE and an exon were selected. We performed in silico translation of these new nucleotide sequences: in the three open reading frames (ORFs) overlapping the breakpoint in case the TE was the donor, or beginning with the gene ORF in case the exon was the donor. This pipeline has generated several files:

1) FASTA sequences of the junctions: these sequences represent the actual novel peptides that could potentially serve as neoantigens. They are essential for downstream analyses such as predicting MHC binding affinity.
2) Detailed JET information table: this table comprises critical information for each JET such as:
    a. TE and gene involved, which could allow for the identification of patterns of preferential involvement of certain TEs or genes in JET formation.
    b. The position of the non-canonical junction, important for understanding the structural implications of these junctions.
    c. TEs acting as a splice donor or acceptor.
    d. The family and superfamily of the TE, which helps in identifying trends in which type of TEs are more prone to forming JETs upon PIN1 ablation.
    e. The number of reads supporting that JET, which provides confidence in the detected JET and allows prioritization of high confidence candidates for further studies.

Comparing the JETs we found that upon siPIN1 there was an increase in their number and in their expression (Figure 16A, 16B). This result suggests that PIN1 may play a role in maintaining genomic stability or regulating splicing events that suppress JETs formation under normal conditions. The increase in JETs could enhance neoantigen load, alter protein function or increase genomic instability.

Since we knew from literature that SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1) is involved in H3K9 methylation, heterochromatin formation and plays a crucial role in TE repression and JETs formation, we wanted to verify the similarities between our RNAseq with PIN1 ablation and the RNAseq from the group of Liu et al. done on MCF7 treated with SETDB1 shRNA (Figure 16C). We found that 34 JETs were common between the two experiments, suggesting a role of PIN1 as an epigenetic regulator of TEs via H3K9me methylation (Figure 16D).

PIN1 inhibition, through its effect on JET formation, might represent a new way to enhance tumor immunogenicity. Based on these results, we want to validate if the produced peptides and neoantigens were immunogenic in a living organism. To accomplish this, we designed an in-vivo experiment utilizing a mouse model. We decided to inject subcutaneously PIN1WT and PIN1KO TNBC 4T1 mouse cell line in mouse model, allowing tumor and TME formation and performing a RNAseq analysis. Intriguingly, our analysis of the mouse model data recapitulated many of the key findings we had previously observed in the human MCF7 BC cell line. We noted an upregulation of pathway linked to IFN/antigen presentation, a de-regulation of TEs expression and an increase in JETs number and expression (Figure 16E-H).

To delve deeper into the potential immunogenicity of the JET-derived peptides, we submitted the FASTA of the JET-derived neopeptides for each sample to the prediction MHC binding affinity (netMHCpan) (Reynisson et al.) tool. This tool predicts binding of peptides to any MHC molecule of known sequence using artificial neural networks (ANNs). As recommended in the Burbage et al. paper, we instructed the tool to analyze peptides of lengths 8, 9, 10, and 11 amino acids. This range of lengths covers the most common sizes of peptides that can be effectively presented by MHC molecules and recognized by T cells. Another necessary input of this tool is the type of MHC molecule, that is specific for each organism. In our case, since the RNAseq was done on the BALB/CJ mouse strain, we select the molecules H-2-Kd,H-2-Dd,H-2-Ld. The analysis revealed several putative neoantigens with different MHC binding affinities. This affinity can be divided into 3 categories based on the concentration of neoantigens needed to be recognized: strong, moderate and low.

After 10 days, splenic CD8+ T-lymphocytes were isolated and restimulated in vitro by exposing them to dendritic cells loaded with synthetic peptides derived from the JETs derived from the RNAseq analysis: 9 PIN1 KO unique peptides with strong and moderate expression levels and high MHC-I affinity, PIN1WT exclusive peptides, poorly expressed peptides, peptides shared between PIN1WT and KO, and weak MHC-I affinity peptides were used (Table 3). In total we selected 14 JETs and the unrelated peptide derived from chicken ovalbumin SIINFEKL. Subsequently, interferon-g secretion by activated CD8+ T-lymphocytes was measured by counting enzyme-linked immunospots (ELISpot). This highly sensitive technique enabled us to detect and quantify cytokine-producing cells at the single-cell level, providing crucial information about the immune response to tumors. We found that only the peptides derived from the PIN1KO cells and

with a strong or moderate affinity can trigger an effective cytotoxic CD8+ T-cell immune response (Figure 17, P1-P9) and vice versa, P10-P15 peptides did not elicit a strong CD8+ cells' activation. SIINFEKL did not activate any immune response. This finding suggests that, in addition to a broader activation of the innate immunity through the pathway of cGAS/STING1 and IFN within the cells, PIN1 ablation could also generate putative neoantigens through reactivation of TEs and increase tumor immunogenicity.

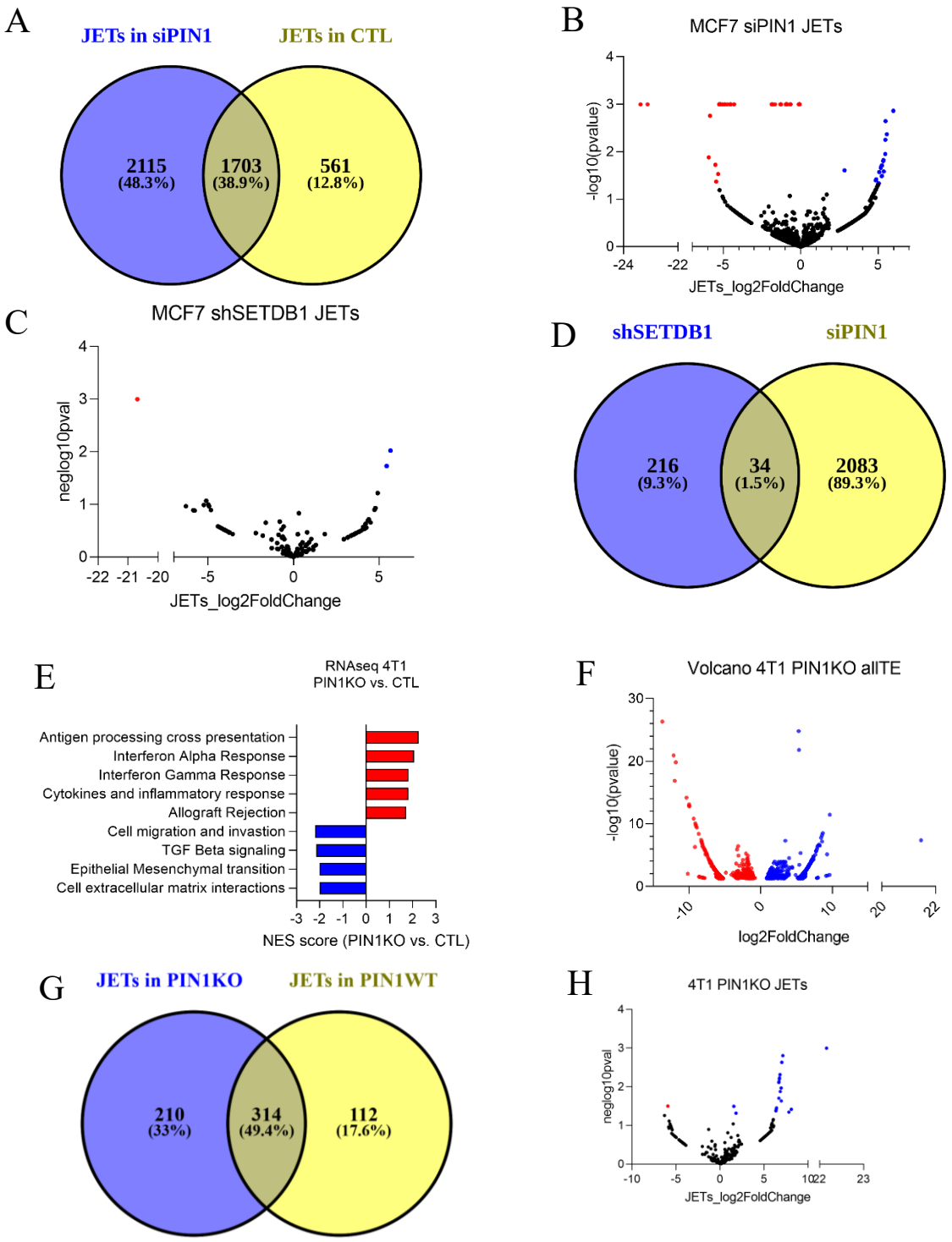

Figure 16: *PIN1 ablation generates alternative splicing between exons and TEs***.** A, B) Venn diagram of JETs retrieved from the previous RNAseq of MCF7 treated with siPIN1 and relative volcano plot of the expressed JETs, showing on the y axis the negative log10 of the pvalue and on the x axis the log2 of the FC (Fold Change). C) Volcano plot of the JETs expressed upon shSETDB1, showing on the y axis the negative log10 of the pvalue and on the x axis the log2 of the FC (Fold Change). D) Venn diagram of the JETs expressed only upon siPIN1 or shSETDB1, showing 34 JETs in common. E) PreRanked GSEA of the DEGs coming from 4T1 PIN1KO calculated using the R package DESEq2: on the y axis there are the categories while on the x axis the NES score in

the comparison PIN1KO vs CTL. F) Volcano plot of the expressed TEs, showing on the y axis the negative log10 of the pvalue and on the x axis the log2 of the FC (Fold Change). G) Venn diagram of the JETs expressed upon PIN1KO or in PIN1WT condition, showing an increase after treatment. H) Volcano plot of the expressed JETs, showing on the y axis the negative log10 of the pvalue and on the x axis the log2 of the FC (Fold Change).

Peptide number	Peptide sequence	Mean host gene ln mRNA expression	Host gene	MHC affinity	Specificity
1	**ASLWCMTSLTRGPRLLFPALT**	9.25	Rab1b	strong	KO
2	**LPECSYWCFAAHVSLRQGC**	9.21	H3f3a	moderate	KO
3	**CFGGVVVVYQGPRWHLLHRYP**	8.64	Tyms	moderate	KO
4	**KSRPFLNNIADFSLSVFPVPD**	8.22	Mpzl3	moderate	KO
5	**SVSQHLGGRGRWPWALALCLC**	8.36	Cox15	moderate	KO
6	**SQHLGGRGRWPWALALCLC**	8.36	Cox15	moderate	KO
7	**SSSMAPSSYSTLRRQRQENLC**	8.23	Trpv4	moderate	KO
8	**EIEAIPQQHSRFFTISLSSTR**	8.60	Mpzl3	moderate	KO
9	**IPALGHWPLDQPPVP**	7.65	Abhd16a	strong	KO
10	**DPCPFLIPSSRVQPLKKRKKL**	10.41	Hmgn1	weak	KO
11	**SIDSYQDGQQSI**	7.14	Pknox1	strong	WT
12	**QSLGGRDRRMPSKIRINLEE**	7.82	E2f6	weak	KO+WT
13	**HHQGGRPGSSSFFPLPL**	5.61	Rad51b	weak	KO
14	**MSLRPAWYYKVLKTSRLNYR**	2.51	Ing4	weak	KO+WT

Table 3: *List of neopeptides discovered using the netMHCpan tool with their aminoacidic sequence, the mean expression of the host gene, its name, the MHC affinity level and their specificity.*

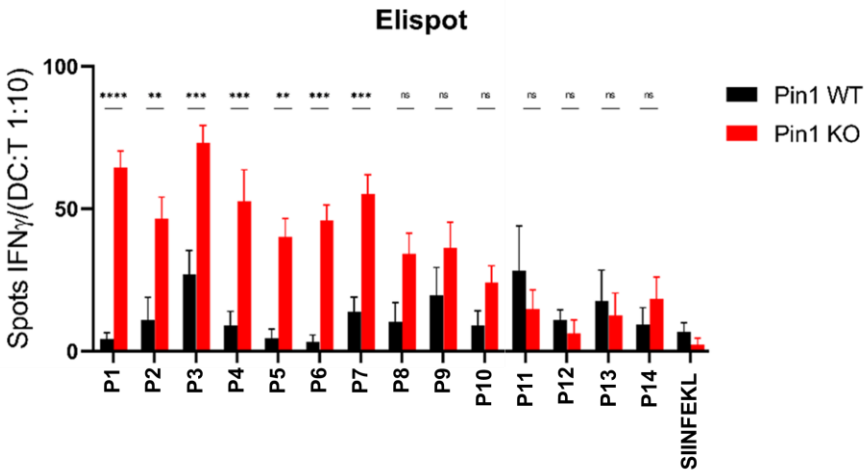

Figure 17: *PIN1 ablation TEs-derived peptides trigger an effective cytotoxic CD8+ T-cell immune response.* Barplot of the selected 15 peptides showing the spots of IFN gamma production in PIN1 WT or PIN1 KO cells. All data are presented as mean ± standard error of the mean; *p<0.1, **p<0.5, ***p<0.01, ****p<0.05. Courtesy of Alessandra Rustighi, Anna Renzi and Lucía López.

3.7 scRNAseq from cPIN1KO in PyMT Mouse Model Shows Increased Immune Infiltrate and Activation of Both Innate and Adaptive Immune Response

Since siPIN1 generated a strong transcriptomic change into tumoral cells, we wanted to address if these changes impact also on the TME composition. Thus, we decided to perform a scRNAseq in a mouse model harboring MMTV-PyMT BC cells with tamoxifen-inducible conditional PIN1KO. We chose to perform this analysis using 10xGenomics technology and we generated four samples, two with PIN1WT and two with PIN1KO, with 10000 expected number of cells each. A total of 32,882 cells were sequenced (15,128 from WT tumors and 17754 from PIN1KO tumors). We then proceeded to perform the next analyses using the R package Seurat (Hao et al.) (version 4). The first step is to filter and normalize the data and we removed outlier cells that contained more than 8,000 genes and 50000 molecules as they could potentially constitute cell doublets. Moreover, we removed cells with less than 1800 genes and 2500 molecules or more than 12% of mitochondrial genes as they could be empty droplets or dying cells. As a result, 28,796 high quality cells (12720 PyMT/WT and 16076 PyMT/PIN1KO) with a total of 28796 genes proceeded with the downstream analyses. We chose to use the *FindVariableFeatures* method using 2000 genes and *RunUMAP*, *FindNeighbors* and *FindClusters* using the first 20 principal components. Next, since we needed a reference to assign each single cell to its correct cell type, we have done extensive literature research and found several cell type markers (Figure 18A) and we used them to perform the clustering and observe how the cells from my data spatially dispose on a UMAP plot. We selected the 0.2 resolution and divided cells into three major cell types (Epithelial, Stromal and Immune) and subset the original Seurat object based on this classification. For each of them, we run again the *RunPCA*, *RunUMAP*, *FindNeighbors*, *FindClusters* and *FindNeighbors* functions, selected the adequate resolution and assigned at every cluster the correct cell type. We obtained 23 cell types shown in Figure 18B and observed that, upon PIN1KO there is a strong reduction in the pro-metastatic basal, LP (luminal progenitor) expressing Krt14 and CAFs population while another population of LP cells expressing a more differentiated marker (Krt19) increased. We then calculated the pseudobulk expression, which is a technique that aggregates gene expression information from individual cells into larger groups, calculating its sum. It serves to make comparisons between different experimental conditions like in this case between the cPIN1KO and the control for each cell type. This technique improves signal-to-noise ratio compared to

individual cell analysis and it allows for the use of well-established bulk RNAseq analysis tools like DESeq2 and GSEA. After having calculated the matrices and the list of genes ordered by their differential expression statistics from DESeq2 for each cell type, we performed a GSEA and looked for the most deregulated pathways related to immunity in the LP population and found an upregulation of pathways linked to immune activity and a downregulation of pathways regarding tumor development and mechano signaling (Figure 18C).

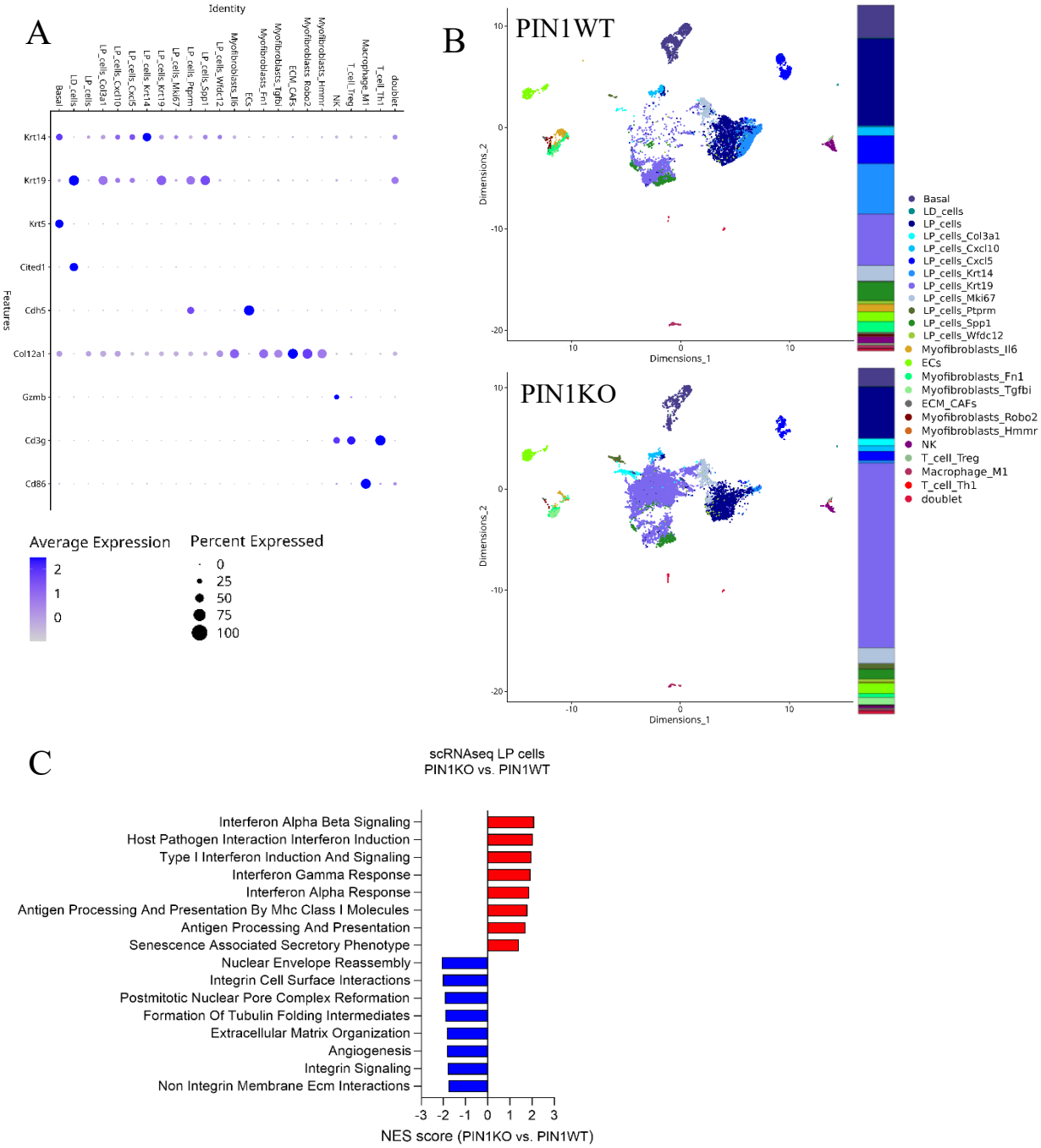

Figure 18: *PIN1KO changes the TME and deregulates tumors survival fundamental pathways.* A) Dot Plot of known specific marker genes of the eight more represented cell populations. B) Uniform Manifold Approximation and Projection (UMAP) and barplot representation of scRNA-seq from PIN1KO and PIN1WT mice samples (PIN1WT: 12720 cells, PIN1KO: 16076 cells) colored according to cell type, divided by sample type. Abbreviations: LP, luminal progenitors; LD luminal differentiated; NK natural killer. C) Gene sets enrichment analysis (GSEA) of immune, tumor development and mechano signaling significantly enriched (pval<0.05) pathways in mammary epithelial cells from PIN1KO vs PIN1WT mouse. The positive normalized enrichment score (NES) indicates the degree to which gene sets are overrepresented in the above-described conditions.

Altogether these results suggest that PIN1 activity is supported by a stiff ECM and maintains an immunosuppressive TME in BC and its inhibition may reduce tumor growth by reprogramming tumor cells and the TME, enhancing immune cell infiltration and activity (Figure 19).

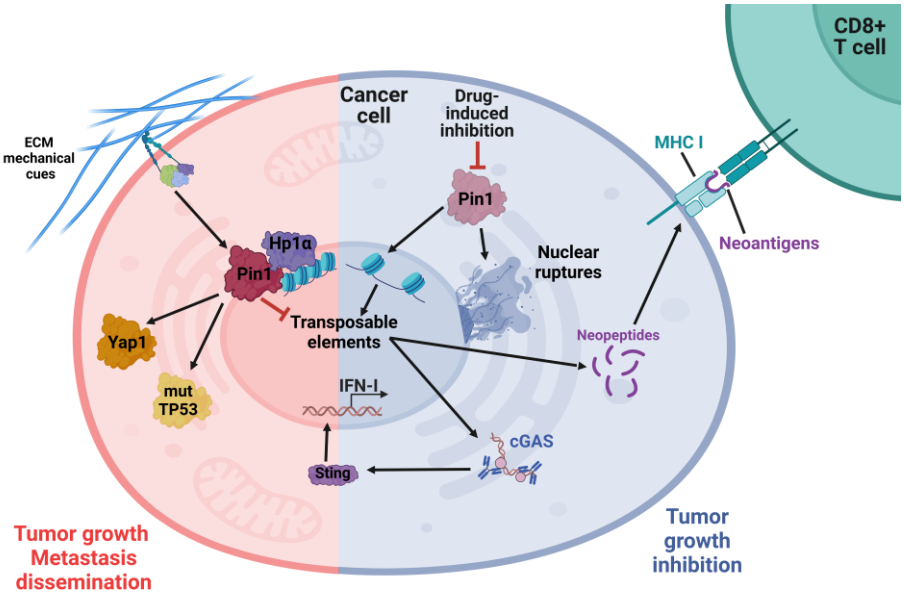

Figure 19: *Schematic model of the impact of PIN1 on tumor growth and the effect of its inhibition on tumor microenvironment. Created in https://BioRender.com*

4. MATERIAL AND METHODS

Cell lines

MCF7 and MCF10A.DCIS.COM are human breast cancer cell lines. MCF7 were cultured in Eagle's Minimum Essential Medium (EMEM, Sigma) supplemented with Fetal Bovine Serum (FBS), 100 U/mL penicillin, 10 μg/mL streptomycin, 1% Minimum essential medium non-essential amino acids (MEM NEAA), and 10 μg/ml recombinant human insulin. MCF-10A.DCIS.COM cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM)/F12.

4T1 is a mouse breast cancer cell line cultured in RPMI-1640 medium supplemented with 10% Fetal Bovine Serum (FBS), 100 U/mL penicillin, 10 μg/mL streptomycin. Human and murine cell lines are from ATCC or other laboratories cooperating on the project. Cells were tested for mycoplasma contamination with negative results.

Transfections

siRNA transfections were performed with Lipofectamine RNAi-MAX (Life technologies) in an antibiotic-free medium, according to manufacturer instructions. As a negative control siRNA, the Qiagen AllStars Negative Control was used. Sequences of siRNAs are reported below.

Oligonucleotide	Sequence
siPIN1#1	GCCAUUUGAAGACGCCUCG
siPIN1#7	CGGGAGAGGAGGACUUUGA
siCapZ	GAAGUACGCUGAACGAGAU

Antibodies

The following antibodies and working concentrations were used for immunofluorescence analysis: anti-Pin1 (1:100, Santa Cruz Biotechnologies sc-46660), anti-YAP1 (1:200, Abcam ab52771), and anti-Lamin B1 (1:300, Abcam ab16048).

Immunofluorescence analysis of mammalian cells

Briefly, cells were fixed in 4% paraformaldehyde for 20 min, washed in phosphate-buffered saline (PBS), permeabilized with 0.1% Triton X-100 for 10 min, and blocked in 3% Fetal Bovine Serum FBS/PBS for 30 min. Antigen recognition was done by incubation with primary antibody at 4 °C

for 14 h and with secondary antibodies (goat anti-mouse Alexa Fluor 568 and goat anti-rabbit Alexa Fluor 488, Life Technologies) at 4 °C for 1 h. Nuclei were stained with Hoechst (Life Technologies, 33342) or Dapi (Sigma Aldrich, 32670-F) for 15 min.

Preparation of fibronectin-coated hydrogel matrix

8 or 0.5kPa Easy Coat hydrogels (Cell guidance system) were coated with 10 μg/ml fibronectin.

Single cell RNA sequencing (scRNAseq) library preparation from PIN1KO and PIN1WT mice

Mammary glands from 8-week-old mice were isolated and digested as previously described. scRNAseq libraries were prepared by using Chromium Next GEM Single Cell 3' Reagent Kits v3.1 Dual Index (10X Genomics) following manufacturer's instructions. An estimate of 16,500 cells were loaded to each channel with the average recovery rate of 10,000 cells. Briefly, cells were resuspended in the master mix and loaded together with partitioning oil and gel beads into the chip to generate the gel bead-in-emulsion (GEM). The poly-A RNA from the cell lysate contained in every single GEM was retrotranscribed to cDNA, which contains an Illumina R1 primer sequence, Unique Molecular Identifier (UMI) and the 10x Barcode. The pooled bar- coded cDNA was then cleaned up with Silane DynaBeads, amplified by PCR and the appropriate size fragments were selected with SPRIselect reagent for subsequent library construction. During the library construction Illumina R2 primer sequence, paired-end con- structs with P5 and P7 sequences and a sample index were added. The final indexed libraries were quantified by using the Cell-free DNA Screen Tape Assay (Agilent) on a TapeStation 4150 (Agilent) and were sequenced on HiSeqX (Illumina).

5. DISCUSSION AND FUTURE DIRECTIONS

5.1 Relationship of PIN1 activity with ECM stiffness and the immune TME

A fundamental challenge in cancer biology lies in understanding and predicting tumor evolution, considering not only the heterogeneity of cancer cells but also TME dynamics. In cancer cells, alterations in oncogenes and tumor suppressor genes play crucial roles in reshaping the TME, thus promoting tumor progression and diminishing therapeutic efficacy (Capaci et al.; Gál et al.; Anderson and Simon; Soysal et al.). For instance, aggressive tumors are characterized by increased collagen density and ECM stiffness and frequently display - poor infiltration and functionality of cytotoxic T-cells, thereby compromising therapy outcomes (Nicolas-Boluda et al.; Kuczek et al.). A stiff ECM impacts also cancer cells themselves, through mechanotransduction pathways that reinforce therapy resistance acting on oncogenes, such as mutant TP53 and YAP1, two well-known mechanically-induced drivers of tumor progression. Of note, both have been shown to be regulated by PIN1, yet the involvement of Pin1 in mechano-signaling only recently has begun to be elucidated. Indeed, work from our group has unveiled PIN1 as a critical regulator of nuclear Lamin B structure and HP1a anchoring, particularly in response to mechanical stress (Napoletano et al.).

Now, work performed in this thesis has contributed to demonstrate that PIN1 activity correlates with ECM stiffness in BC, in association with immunosuppressive milieus and poor immune cell infiltration. Furthermore, experiments performed in human and mouse preclinical BC models allowed us to unveil that PIN1 levels and activity are upregulated by mechanical cues, and that PIN1 acts as a critical regulator of immune surveillance. Using immune-cold mammary tumor models (MMTV-PyMT, 4T1-Balb/c), we demonstrated that ablating PIN1 in cancer cells suppresses tumor and metastasis growth by altering nuclear and heterochromatin structure, thus activating TEs, thus stimulating a cGAS/STING/IFN-I response and formation of JET-derived peptides, eliciting a potent anti-tumor immune response. Thus, by modulating the composition of the TME, PIN1 promotes an immunosuppressive environment that enhances tumor survival and resistance to treatment. This discovery not only advances our understanding of the role of the TME in immune evasion, but also positions PIN1 as a potential target for therapeutic strategies aimed at altering the immunological landscape of tumors to improve patient responses to cancer immunotherapies.

5.2 Therapeutic opportunities

Prompted by our results, an interesting opportunity will be to test whether Pin1 inhibition could impact the growth and response to therapy of aggressive and immune-cold tumors. To this aim FDA approved drugs shown to perform as specific PIN1 inhibitors, such as all-trans retinoic acid (ATRA) and arsenic Trioxide (ATO) could be employed, alone and in combination with standard BC therapies, to assess whether they could sensitize to ICB, in mice bearing immune cold BC primary tumors and metastases.

In addition, PIN1 inhibitors could synergize with molecules that impact nuclear and heterochromatin integrity, such as CDK4/6-inhibitors (Palbociclib, Abemaciclib). Interestingly, recent evidence has demonstrated that CDK4/6-inhibitors (Palbociclib, Abemaciclib), already used in combination with ICB and endocrine therapy in a phase I/II trial for patients with HR+/HER2- locally advanced or metastatic BC synergize with PIN1 inhibition to improve ICB outcome in BC mouse models (Ke et al.; LeVee et al.). Another therapeutic opportunity is given by the use of inhibitors of ECM stiffness and mechanotransduction (Tschumperlin and Lagares), such as FDA-approved FAK-inhibitors in combination with PIN1 inhibition. Also, STING agonists, which are under intense scrutiny as stimulators of anti-tumor immunity, could be used in combination with PIN1 inhibitors, which may provide promising results in enhancing the efficacy of existing immunotherapies (Wang et al.).

5.3 The Challenge of Breast Cancers Subtypes

The PAM50 gene signature enables classification of BCs into five intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like. This classification system has become a foundational element in breast cancer research, significantly advancing our understanding of tumor biology and guiding the development of subtype-specific therapeutic approaches. By aligning treatment strategies with these well-defined subtypes, the PAM50 classification has helped personalize cancer care and improve outcomes for many patients. However, despite its substantial impact, the accuracy of this classification system can be limited by intertumoral heterogeneity, which persists even within the boundaries of these established subtypes (Ohara et al.; Raj-Kumar et al.).

Of note, PIN1 overexpression has been observed in all BC subtypes and correlates with the expression of subtype-specific molecular markers, such as ERα in Luminal subtype, ERBB2 in

HER2 subtype, and also with mutant p53 in TNBC subtype (Rustighi et al.). Also, PIN1 levels were shown to display prognostic value in combination with some of these markers, and also with mutant p53 protein levels (Girardini et al.). This evidence implies that PIN1 promotes tumor growth and could be targeted across all BC subtypes. Based on our findings, PIN1 activity signature could be used, also in combination with signatures predictive of immune response (e.g., TILs, T-cells and MHC-II signatures), to improve stratification of patients and predict treatment response for each BC subtype.

In our analysis of transcriptomic data derived from hundreds of BC patients, we encountered limitations related to sample size within each PAM50-defined subtype. The restricted number of samples in certain subtypes hindered our ability to pursue a more granular subclassification that could better reflect underlying biological diversity. Future investigations that incorporate high-resolution single-cell data and subtype-specific analyses are essential to elucidate the impact of PIN1 in each subtype. Furthermore, based on our findings that PIN1 controls heterochromatin dynamics and thus TEs activity we aim to leverage sophisticated analytical tools to investigate potential correlations between PAM50-defined subtypes (or more refined subclassifications) and TEs regulation.

5.4 Advances and limitations in sequencing resolution for the analysis of the impact of PIN1 in the TME

The advent of high-throughput sequencing methodologies, including scRNA-seq and spatial transcriptomics, has revolutionized the characterization of tumors, providing unprecedented insights into the heterogeneity of the TME. Nevertheless, these technological advancements are accompanied by distinct technical limitations. scRNA-seq data is characterized by substantial dropout events, resulting in expression matrices with pervasive zero values that may reflect either genuine biological absence of expression or technical artifacts, presenting challenges in accurately delineating PIN1 expression levels across various TME regions. The extent of this sparsity varies with both the specific scRNA-seq platform employed and sequencing depth (Lähnemann et al.). These dropout issues necessitate careful consideration and often require the application of imputation techniques or specialized bioinformatic pipelines to distinguish genuine biological signals from noise.

Spatial transcriptomics, while offering crucial positional context, relies on probe-based detection methods reminiscent of traditional microarray technology, rather than sequencing-based approaches. This reliance on pre-designed probes limits detection to a subset of predefined transcripts, which constrains the resolution of gene expression profiles and may exclude critical transcripts not included in the probe set, such as all the non-coding part of the genome including TEs, which we have seen to be important in the study of tumors and their treatment or the proteomic alterations instigated by PIN1 in response to ECM stiffness. Despite these limitations, spatial transcriptomics offers a unique advantage over scRNA-seq in that it can capture the spatial organization of the tumor microenvironment (TME). This spatial dimension is indispensable for investigating the heterogeneity of the TME and characterizing distinct cellular neighborhoods within tumor tissues, which are often associated with differential therapeutic responses and immune evasion mechanisms.

5.5 Artificial Intelligence (AI) and Omics approaches in cancer research

The integration of AI with omics methodologies demonstrates significant potential for managing intricate datasets, elucidating prospective biomarkers, and forecasting patient responses. Nevertheless, concerns have raised regarding an excessive dependence on AI, as the models may be adversely affected by biases stemming from data quality and accessibility, which could result in overfitting, particularly when confronted with data scarcity in scRNA-seq datasets. Furthermore, the integration of omics data presents substantial difficulties due to the varying stability and technical variability inherent in distinct omics types (e.g., lipidomics, transcriptomics, and proteomics), which may compromise the accuracy of AI models.

In contrast to the assertion that AI will effortlessly transform precision medicine, skepticism exists regarding the capacity of contemporary AI models to adequately encapsulate the intricate dynamics of TME interactions, especially within the context of heterogeneous PIN1-expressing tumors. A vital factor to consider is that while AI possesses the capability to discern patterns, it may overlook context-dependent regulatory mechanisms, such as the role of PIN1 in immune modulation and ECM stiffness modulation. The integration of interpretable AI, rather than an exclusive focus on predictive models, could facilitate the contextualization of findings within the broader framework of TME interactions, thereby illuminating both synergistic and antagonistic functions of PIN1.

6. ACKNOWLEDGEMENTS

I would like to thank all the people who have accompanied me, from a professional and personal point of view, during my PhD. First, I would like to thank my supervisor Prof. Giannino Del Sal that three years ago gave me the opportunity to join his lab and to carry out this interesting research project at the University of Trieste and at the International Centre for Genetic Engineering and Biotechnology (ICGEB, Trieste). Thanks for the precious scientific discussions and scientific advice about my research activity. I would like to thank also all the members of the Cancer Cell Signaling group for their collaboration and support, in particular Alessandra Rustighi, Rebecca Bertolio and Francesco Napoletano. A special thanks goes to Dr. Silvano Piazza who carefully guided, followed and trained me on this project during these years. Thanks also for teaching me his knowledge and for the fundamental contribution he gave in writing and correcting this thesis. I would like to thank Prof. Mattia Forcato and Prof. Claudio Tripodo that kindly have read my thesis. A particular thanks goes to all my current and past colleagues for giving me help and support me when I needed it. I want to thank my girlfriend Alessandra and my friend Emanuele, who have never left me alone in the most difficult moments and who have brightened my days during these years. A deep thanks goes to my parents and my family who have always given me the strength, for their constant, fundamental and precious support. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

REFERENCES

Akhoon, Neha. ‘Precision Medicine: A New Paradigm in Therapeutics’. *International Journal of Preventive Medicine*, vol. 12, 2021, p. 12. *PubMed*, https://doi.org/10.4103/ijpvm.IJPVM_375_19.

Alvarado-Ortiz, Eduardo, et al. ‘Mutant P53 Gain-of-Function: Role in Cancer Development, Progression, and Therapeutic Approaches’. *Frontiers in Cell and Developmental Biology*, vol. 8, 2020, p. 607670. *PubMed*, https://doi.org/10.3389/fcell.2020.607670.

Anand, Uttpal, et al. ‘Cancer Chemotherapy and beyond: Current Status, Drug Candidates, Associated Risks and Progress in Targeted Therapeutics’. *Genes & Diseases*, vol. 10, no. 4, July 2023, pp. 1367–401. *PubMed*, https://doi.org/10.1016/j.gendis.2022.02.007.

Anderson, Nicole M., and M. Celeste Simon. ‘The Tumor Microenvironment’. *Current Biology: CB*, vol. 30, no. 16, Aug. 2020, pp. R921–25. *PubMed*, https://doi.org/10.1016/j.cub.2020.06.081.

Bagaev, Alexander, et al. ‘Conserved Pan-Cancer Microenvironment Subtypes Predict Response to Immunotherapy’. *Cancer Cell*, vol. 39, no. 6, June 2021, pp. 845-865.e7. *PubMed*, https://doi.org/10.1016/j.ccell.2021.04.014.

Ballot, Elise, et al. ‘Tumor Infiltrating Lymphocytes Signature as a New Pan-Cancer Predictive Biomarker of Anti PD-1/PD-L1 Efficacy’. *Cancers*, vol. 12, no. 9, Aug. 2020, p. 2418. *PubMed*, https://doi.org/10.3390/cancers12092418.

Bendall, Matthew L., et al. ‘Telescope: Characterization of the Retrotranscriptome by Accurate Estimation of Transposable Element Expression’. *PLoS Computational Biology*, vol. 15, no. 9, Sept. 2019, p. e1006453. *PubMed*, https://doi.org/10.1371/journal.pcbi.1006453.

Berke, Taylor P., et al. 'Role of Reactivating Mutant P53 Protein in Suppressing Growth and Metastasis of Triple-Negative Breast Cancer'. *OncoTargets and Therapy*, vol. 15, 2022, pp. 23–30. *PubMed*, https://doi.org/10.2147/OTT.S342292.

Bertolio, Rebecca, et al. 'Dynamic Links between Mechanical Forces and Metabolism Shape the Tumor Milieu'. *Current Opinion in Cell Biology*, vol. 84, Oct. 2023, p. 102218. *PubMed*, https://doi.org/10.1016/j.ceb.2023.102218.

Boutelle, Anthony M., and Laura D. Attardi. 'P53 and Tumor Suppression: It Takes a Network'. *Trends in Cell Biology*, vol. 31, no. 4, Apr. 2021, pp. 298–310. *PubMed*, https://doi.org/10.1016/j.tcb.2020.12.011.

Burbage, Marianne, et al. 'Epigenetically Controlled Tumor Antigens Derived from Splice Junctions between Exons and Transposable Elements'. *Science Immunology*, vol. 8, no. 80, Feb. 2023, p. eabm6360. *DOI.org (Crossref)*, https://doi.org/10.1126/sciimmunol.abm6360.

Campaner, Elena, et al. 'A Covalent PIN1 Inhibitor Selectively Targets Cancer Cells by a Dual Mechanism of Action'. *Nature Communications*, vol. 8, June 2017, p. 15772. *PubMed*, https://doi.org/10.1038/ncomms15772.

Capaci, Valeria, et al. 'Amplifying Tumor-Stroma Communication: An Emerging Oncogenic Function of Mutant P53'. *Frontiers in Oncology*, vol. 10, 2020, p. 614230. *PubMed*, https://doi.org/10.3389/fonc.2020.614230.

Chen, Yang, et al. 'Prolyl Isomerase Pin1: A Promoter of Cancer and a Target for Therapy'. *Cell Death & Disease*, vol. 9, no. 9, Aug. 2018, p. 883. *PubMed*, https://doi.org/10.1038/s41419-018-0844-y.

Chow, Melvyn T., et al. ‘Inflammation and Immune Surveillance in Cancer’. *Seminars in Cancer Biology*, vol. 22, no. 1, Feb. 2012, pp. 23–32. *PubMed*, https://doi.org/10.1016/j.semcancer.2011.12.004.

Chuang, Hsiang-Hao, et al. ‘Targeting Pin1 for Modulation of Cell Motility and Cancer Therapy’. *Biomedicines*, vol. 9, no. 4, Mar. 2021, p. 359. *PubMed*, https://doi.org/10.3390/biomedicines9040359.

Cunningham, Fiona, et al. ‘Ensembl 2022’. *Nucleic Acids Research*, vol. 50, no. D1, Jan. 2022, pp. D988–95. *PubMed*, https://doi.org/10.1093/nar/gkab1049.

Curtis, Christina, et al. ‘The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups’. *Nature*, vol. 486, no. 7403, Apr. 2012, pp. 346–52. *PubMed*, https://doi.org/10.1038/nature10983.

Dank, Magdolna, et al. ‘The Prediction Analysis of Microarray 50 (PAM50) Gene Expression Classifier Utilized in Indeterminate-Risk Breast Cancer Patients in Hungary: A Consecutive 5-Year Experience’. *Genes*, vol. 14, no. 9, Aug. 2023, p. 1708. *PubMed*, https://doi.org/10.3390/genes14091708.

de Anda-Jáuregui, Guillermo, and Enrique Hernández-Lemus. ‘Computational Oncology in the Multi-Omics Era: State of the Art’. *Frontiers in Oncology*, vol. 10, 2020, p. 423. *PubMed*, https://doi.org/10.3389/fonc.2020.00423.

Deepak, K. G. K., et al. ‘Tumor Microenvironment: Challenges and Opportunities in Targeting Metastasis of Triple Negative Breast Cancer’. *Pharmacological Research*, vol. 153, Mar. 2020, p. 104683. *PubMed*, https://doi.org/10.1016/j.phrs.2020.104683.

Dobin, Alexander, et al. ‘STAR: Ultrafast Universal RNA-Seq Aligner’. *Bioinformatics (Oxford, England)*, vol. 29, no. 1, Jan. 2013, pp. 15–21. *PubMed*, https://doi.org/10.1093/bioinformatics/bts635.

Emon, Bashar, et al. ‘Biophysics of Tumor Microenvironment and Cancer Metastasis - A Mini Review’. *Computational and Structural Biotechnology Journal*, vol. 16, 2018, pp. 279–87. *PubMed*, https://doi.org/10.1016/j.csbj.2018.07.003.

Foroutan, Momeneh, et al. ‘Single Sample Scoring of Molecular Phenotypes’. *BMC Bioinformatics*, vol. 19, no. 1, Dec. 2018, p. 404. *DOI.org (Crossref)*, https://doi.org/10.1186/s12859-018-2435-4.

Gál, Peter, et al. ‘How Signaling Molecules Regulate Tumor Microenvironment: Parallels to Wound Repair’. *Molecules (Basel, Switzerland)*, vol. 22, no. 11, Oct. 2017, p. 1818. *PubMed*, https://doi.org/10.3390/molecules22111818.

Giorello, María Belén, et al. ‘Cancer-Associated Fibroblasts in the Breast Tumor Microenvironment’. *Journal of Mammary Gland Biology and Neoplasia*, vol. 26, no. 2, June 2021, pp. 135–55. *PubMed*, https://doi.org/10.1007/s10911-020-09475-y.

Girardini, Javier E., et al. ‘A Pin1/Mutant P53 Axis Promotes Aggressiveness in Breast Cancer’. *Cancer Cell*, vol. 20, no. 1, July 2011, pp. 79–91. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ccr.2011.06.004.

Greco, Luana, et al. ‘Microsatellite Instability and Immune Response: From Microenvironment Features to Therapeutic Actionability-Lessons from Colorectal Cancer’. *Genes*, vol. 14, no. 6, May 2023, p. 1169. *PubMed*, https://doi.org/10.3390/genes14061169.

Grünwald, Barbara T., et al. ‘Spatially Confined Sub-Tumor Microenvironments in Pancreatic Cancer’. *Cell*, vol. 184, no. 22, Oct. 2021, pp. 5577-5592.e18. *PubMed*, https://doi.org/10.1016/j.cell.2021.09.022.

Guo, Liantao, et al. ‘Breast Cancer Heterogeneity and Its Implication in Personalized Precision Therapy’. *Experimental Hematology & Oncology*, vol. 12, no. 1, Jan. 2023, p. 3. *PubMed*, https://doi.org/10.1186/s40164-022-00363-1.

Hao, Yuhan, et al. ‘Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis’. *Nature Biotechnology*, vol. 42, no. 2, Feb. 2024, pp. 293–304. *PubMed*, https://doi.org/10.1038/s41587-023-01767-y.

Harrell Jr, Frank E. *Hmisc: Harrell Miscellaneous*. 17 July 2003, p. 5.2-0. *DOI.org (Crossref)*, https://doi.org/10.32614/CRAN.package.Hmisc.

Hayward, Mary-Kate, et al. ‘Tissue Mechanics in Stem Cell Fate, Development, and Cancer’. *Developmental Cell*, vol. 56, no. 13, July 2021, pp. 1833–47. *PubMed*, https://doi.org/10.1016/j.devcel.2021.05.011.

Heumos, Lukas, et al. ‘Best Practices for Single-Cell Analysis across Modalities’. *Nature Reviews. Genetics*, vol. 24, no. 8, Aug. 2023, pp. 550–72. *PubMed*, https://doi.org/10.1038/s41576-023-00586-w.

Humphrey, Jay D., et al. ‘Mechanotransduction and Extracellular Matrix Homeostasis’. *Nature Reviews. Molecular Cell Biology*, vol. 15, no. 12, Dec. 2014, pp. 802–12. *PubMed*, https://doi.org/10.1038/nrm3896.

Ingallina, Eleonora, et al. ‘Mechanical Cues Control Mutant P53 Stability through a Mevalonate-RhoA Axis’. *Nature Cell Biology*, vol. 20, no. 1, Jan. 2018, pp. 28–35. *PubMed*, https://doi.org/10.1038/s41556-017-0009-8.

Jia, Qingzhu, et al. ‘Heterogeneity of the Tumor Immune Microenvironment and Its Clinical Relevance’. *Experimental Hematology & Oncology*, vol. 11, no. 1, Apr. 2022, p. 24. *PubMed*, https://doi.org/10.1186/s40164-022-00277-y.

Jiang, Yangfu, et al. ‘Targeting Extracellular Matrix Stiffness and Mechanotransducers to Improve Cancer Therapy’. *Journal of Hematology & Oncology*, vol. 15, no. 1, Mar. 2022, p. 34. *PubMed*, https://doi.org/10.1186/s13045-022-01252-0.

Johnson, Douglas B., et al. 'Immune-Checkpoint Inhibitors: Long-Term Implications of Toxicity'. *Nature Reviews. Clinical Oncology*, vol. 19, no. 4, Apr. 2022, pp. 254–67. *PubMed*, https://doi.org/10.1038/s41571-022-00600-w.

Jovic, Dragomirka, et al. 'Single-Cell RNA Sequencing Technologies and Applications: A Brief Overview'. *Clinical and Translational Medicine*, vol. 12, no. 3, Mar. 2022, p. e694. *PubMed*, https://doi.org/10.1002/ctm2.694.

Kandoth, Cyriac, et al. 'Mutational Landscape and Significance across 12 Major Cancer Types'. *Nature*, vol. 502, no. 7471, Oct. 2013, pp. 333–39. *PubMed*, https://doi.org/10.1038/nature12634.

Ke, Shizhong, et al. 'Reciprocal Inhibition of PIN1 and APC/CCDH1 Controls Timely G1/S Transition and Creates Therapeutic Vulnerability'. *Research Square*, Jan. 2023, p. rs.3.rs-2447544. *PubMed*, https://doi.org/10.21203/rs.3.rs-2447544/v1.

Kent, W. James, et al. 'The Human Genome Browser at UCSC'. *Genome Research*, vol. 12, no. 6, June 2002, pp. 996–1006. *PubMed*, https://doi.org/10.1101/gr.229102.

Kong, Yu, et al. 'Transposable Element Expression in Tumors Is Associated with Immune Infiltration and Increased Antigenicity'. *Nature Communications*, vol. 10, no. 1, Nov. 2019, p. 5228. *PubMed*, https://doi.org/10.1038/s41467-019-13035-2.

Koutsogiannouli, Evangelia, et al. 'Complexity in Cancer Biology: Is Systems Biology the Answer?' *Cancer Medicine*, vol. 2, no. 2, Apr. 2013, pp. 164–77. *PubMed*, https://doi.org/10.1002/cam4.62.

Kuczek, Dorota E., et al. 'Collagen Density Regulates the Activity of Tumor-Infiltrating T Cells'. *Journal for Immunotherapy of Cancer*, vol. 7, no. 1, Mar. 2019, p. 68. *PubMed*, https://doi.org/10.1186/s40425-019-0556-6.

Kwon, John, and Samuel F. Bakhoum. 'The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer'. *Cancer Discovery*, vol. 10, no. 1, Jan. 2020, pp. 26–39. *PubMed*, https://doi.org/10.1158/2159-8290.CD-19-0761.

Lähnemann, David, et al. 'Eleven Grand Challenges in Single-Cell Data Science'. *Genome Biology*, vol. 21, no. 1, Feb. 2020, p. 31. *PubMed*, https://doi.org/10.1186/s13059-020-1926-6.

Lanciano, Sophie, and Gael Cristofari. 'Measuring and Interpreting Transposable Element Expression'. *Nature Reviews. Genetics*, vol. 21, no. 12, Dec. 2020, pp. 721–36. *PubMed*, https://doi.org/10.1038/s41576-020-0251-y.

Lander, E. S., et al. 'Initial Sequencing and Analysis of the Human Genome'. *Nature*, vol. 409, no. 6822, Feb. 2001, pp. 860–921. *PubMed*, https://doi.org/10.1038/35057062.

Larsen, Anne Mette H., et al. 'Collagen Density Modulates the Immunosuppressive Functions of Macrophages'. *Journal of Immunology (Baltimore, Md.: 1950)*, vol. 205, no. 5, Sept. 2020, pp. 1461–72. *PubMed*, https://doi.org/10.4049/jimmunol.1900789.

LeVee, Alexis Ann, et al. 'A Phase I/II Trial of Palbociclib, Pembrolizumab, and Endocrine Therapy for Patients with HR+/HER2- Locally Advanced or Metastatic Breast Cancer (MBC): Clinical Outcomes and Stool Microbial Profiling.' *Journal of Clinical Oncology*, vol. 42, no. 16_suppl, June 2024, pp. 1038–1038. *DOI.org (Crossref)*, https://doi.org/10.1200/JCO.2024.42.16_suppl.1038.

Li, Ying, et al. 'Regulation and Mechanism of YAP/TAZ in the Mechanical Microenvironment of Stem Cells (Review)'. *Molecular Medicine Reports*, vol. 24, no. 1, July 2021, p. 506. *PubMed*, https://doi.org/10.3892/mmr.2021.12145.

Liu, Yingqi, et al. ‘Mechanotransduction in Response to ECM Stiffening Impairs cGAS Immune Signaling in Tumor Cells’. *Cell Reports*, vol. 42, no. 10, Oct. 2023, p. 113213. *PubMed*, https://doi.org/10.1016/j.celrep.2023.113213.

López, Lucía, et al. ‘Dendritic Cell-Targeted Therapy Expands CD8 T Cell Responses to Bona-Fide Neoantigens in Lung Tumors’. *Nature Communications*, vol. 15, no. 1, Mar. 2024, p. 2280. *PubMed*, https://doi.org/10.1038/s41467-024-46685-y.

Love, Michael I., et al. ‘Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2’. *Genome Biology*, vol. 15, no. 12, Dec. 2014, p. 550. *DOI.org (Crossref)*, https://doi.org/10.1186/s13059-014-0550-8.

Maller, Ori, et al. ‘Tumour-Associated Macrophages Drive Stromal Cell-Dependent Collagen Crosslinking and Stiffening to Promote Breast Cancer Aggression’. *Nature Materials*, vol. 20, no. 4, Apr. 2021, pp. 548–59. *PubMed*, https://doi.org/10.1038/s41563-020-00849-5.

Mantovani, Fiamma, et al. ‘Mutant P53 as a Guardian of the Cancer Cell’. *Cell Death & Differentiation*, vol. 26, no. 2, Feb. 2019, pp. 199–212. *DOI.org (Crossref)*, https://doi.org/10.1038/s41418-018-0246-9.

Martin, Marcel. ‘Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads’. *EMBnet.Journal*, vol. 17, no. 1, May 2011, p. 10. *DOI.org (Crossref)*, https://doi.org/10.14806/ej.17.1.200.

McKeage, Kate, and Caroline M. Perry. ‘Trastuzumab: A Review of Its Use in the Treatment of Metastatic Breast Cancer Overexpressing HER2’. *Drugs*, vol. 62, no. 1, 2002, pp. 209–43. *PubMed*, https://doi.org/10.2165/00003495-200262010-00008.

Mehraj, Umar, et al. ‘Tumor Microenvironment Promotes Breast Cancer Chemoresistance’. *Cancer Chemotherapy and Pharmacology*, vol. 87, no. 2, Feb. 2021, pp. 147–58. *PubMed*, https://doi.org/10.1007/s00280-020-04222-w.

Merlotti, Antonela, et al. 'Noncanonical Splicing Junctions between Exons and Transposable Elements Represent a Source of Immunogenic Recurrent Neo-Antigens in Patients with Lung Cancer'. *Science Immunology*, vol. 8, no. 80, Feb. 2023, p. eabm6359. *DOI.org (Crossref)*, https://doi.org/10.1126/sciimmunol.abm6359.

Min, Hye-Young, and Ho-Young Lee. 'Molecular Targeted Therapy for Anticancer Treatment'. *Experimental & Molecular Medicine*, vol. 54, no. 10, Oct. 2022, pp. 1670–94. *PubMed*, https://doi.org/10.1038/s12276-022-00864-3.

Mittal, Suruchi, et al. 'The Breast Tumor Microenvironment: Role in Cancer Development, Progression and Response to Therapy'. *Expert Review of Molecular Diagnostics*, vol. 18, no. 3, Mar. 2018, pp. 227–43. *PubMed*, https://doi.org/10.1080/14737159.2018.1439382.

Mohammadi, Hamid, and Erik Sahai. 'Mechanisms and Impact of Altered Tumour Mechanics'. *Nature Cell Biology*, vol. 20, no. 7, July 2018, pp. 766–74. *DOI.org (Crossref)*, https://doi.org/10.1038/s41556-018-0131-2.

Mootha, Vamsi K., et al. 'PGC-1alpha-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes'. *Nature Genetics*, vol. 34, no. 3, July 2003, pp. 267–73. *PubMed*, https://doi.org/10.1038/ng1180.

Napoletano, Francesco, et al. 'The Prolyl-Isomerase PIN1 Is Essential for Nuclear Lamin-B Structure and Function and Protects Heterochromatin under Mechanical Stress'. *Cell Reports*, vol. 36, no. 11, Sept. 2021, p. 109694. *DOI.org (Crossref)*, https://doi.org/10.1016/j.celrep.2021.109694.

Nicolas-Boluda, Alba, et al. 'Tumor Stiffening Reversion through Collagen Crosslinking Inhibition Improves T Cell Migration and Anti-PD-1 Treatment'. *eLife*, vol. 10, June 2021, p. e58688. *PubMed*, https://doi.org/10.7554/eLife.58688.

Northey, Jason J., et al. ‘Tissue Force Programs Cell Fate and Tumor Aggression’. *Cancer Discovery*, vol. 7, no. 11, Nov. 2017, pp. 1224–37. *PubMed*, https://doi.org/10.1158/2159-8290.CD-16-0733.

Ohara, Ako Matsushita, et al. ‘PAM50 for Prediction of Response to Neoadjuvant Chemotherapy for ER-Positive Breast Cancer’. *Breast Cancer Research and Treatment*, vol. 173, no. 3, Feb. 2019, pp. 533–43. *PubMed*, https://doi.org/10.1007/s10549-018-5020-7.

Panciera, Tito, et al. ‘Reprogramming Normal Cells into Tumour Precursors Requires ECM Stiffness and Oncogene-Mediated Changes of Cell Mechanical Properties’. *Nature Materials*, vol. 19, no. 7, July 2020, pp. 797–806. *PubMed*, https://doi.org/10.1038/s41563-020-0615-x.

Park, So-Yeon, and Jeong-Seok Nam. ‘The Force Awakens: Metastatic Dormant Cancer Cells’. *Experimental & Molecular Medicine*, vol. 52, no. 4, Apr. 2020, pp. 569–81. *PubMed*, https://doi.org/10.1038/s12276-020-0423-z.

Piersma, Bram, et al. ‘Fibrosis and Cancer: A Strained Relationship’. *Biochimica Et Biophysica Acta. Reviews on Cancer*, vol. 1873, no. 2, Apr. 2020, p. 188356. *PubMed*, https://doi.org/10.1016/j.bbcan.2020.188356.

Prat, A., et al. ‘PAM50 Assay and the Three-Gene Model for Identifying the Major and Clinically Relevant Molecular Subtypes of Breast Cancer’. *Breast Cancer Research and Treatment*, vol. 135, no. 1, Aug. 2012, pp. 301–06. *PubMed*, https://doi.org/10.1007/s10549-012-2143-0.

Pratt, Stephen J. P., et al. ‘The Mechanical Microenvironment in Breast Cancer’. *Cancers*, vol. 12, no. 6, June 2020, p. 1452. *PubMed*, https://doi.org/10.3390/cancers12061452.

Raj-Kumar, Praveen-Kumar, et al. ‘PCA-PAM50 Improves Consistency between Breast Cancer Intrinsic and Clinical Subtyping Reclassifying a Subset of Luminal A Tumors as Luminal

B'. *Scientific Reports*, vol. 9, no. 1, May 2019, p. 7956. *PubMed*, https://doi.org/10.1038/s41598-019-44339-4.

Rao, Mohan S., et al. 'Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies'. *Frontiers in Genetics*, vol. 9, 2018, p. 636. *PubMed*, https://doi.org/10.3389/fgene.2018.00636.

Reynisson, Birkir, et al. 'NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data'. *Nucleic Acids Research*, vol. 48, no. W1, July 2020, pp. W449–54. *DOI.org (Crossref)*, https://doi.org/10.1093/nar/gkaa379.

Roulot, Aurélie, et al. 'Tumoral Heterogeneity of Breast Cancer'. *Annales De Biologie Clinique*, vol. 74, no. 6, Dec. 2016, pp. 653–60. *PubMed*, https://doi.org/10.1684/abc.2016.1192.

Rustighi, Alessandra, et al. 'Prolyl-Isomerase Pin1 Controls Normal and Cancer Stem Cells of the Breast'. *EMBO Molecular Medicine*, vol. 6, no. 1, Jan. 2014, pp. 99–119. *PubMed*, https://doi.org/10.1002/emmm.201302909.

Sanchez-Vega, Francisco, et al. 'Oncogenic Signaling Pathways in The Cancer Genome Atlas'. *Cell*, vol. 173, no. 2, Apr. 2018, pp. 321-337.e10. *DOI.org (Crossref)*, https://doi.org/10.1016/j.cell.2018.03.035.

Satam, Heena, et al. 'Next-Generation Sequencing Technology: Current Trends and Advancements'. *Biology*, vol. 12, no. 7, July 2023, p. 997. *PubMed*, https://doi.org/10.3390/biology12070997.

Sever, Richard, and Joan S. Brugge. 'Signal Transduction in Cancer'. *Cold Spring Harbor Perspectives in Medicine*, vol. 5, no. 4, Apr. 2015, p. a006098. *PubMed*, https://doi.org/10.1101/cshperspect.a006098.

Singer, Jochen, et al. ‘Bioinformatics for Precision Oncology’. *Briefings in Bioinformatics*, vol. 20, no. 3, May 2019, pp. 778–88. *PubMed*, https://doi.org/10.1093/bib/bbx143.

Soysal, Savas D., et al. ‘Role of the Tumor Microenvironment in Breast Cancer’. *Pathobiology: Journal of Immunopathology, Molecular and Cellular Biology*, vol. 82, no. 3–4, Sept. 2015, pp. 142–52. *PubMed*, https://doi.org/10.1159/000430499.

Stowers, Ryan S., et al. ‘Matrix Stiffness Induces a Tumorigenic Phenotype in Mammary Epithelium through Changes in Chromatin Accessibility’. *Nature Biomedical Engineering*, vol. 3, no. 12, July 2019, pp. 1009–19. *DOI.org (Crossref)*, https://doi.org/10.1038/s41551-019-0420-5.

Subramanian, Aravind, et al. ‘Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles’. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, Oct. 2005, pp. 15545–50. *PubMed*, https://doi.org/10.1073/pnas.0506580102.

Tombari, Camilla, et al. ‘Mutant P53 Sustains Serine-Glycine Synthesis and Essential Amino Acids Intake Promoting Breast Cancer Growth’. *Nature Communications*, vol. 14, no. 1, Oct. 2023, p. 6777. *DOI.org (Crossref)*, https://doi.org/10.1038/s41467-023-42458-1.

Tschumperlin, Daniel J., and David Lagares. ‘Mechano-Therapeutics: Targeting Mechanical Signaling in Fibrosis and Tumor Stroma’. *Pharmacology & Therapeutics*, vol. 212, Aug. 2020, p. 107575. *PubMed*, https://doi.org/10.1016/j.pharmthera.2020.107575.

Vijayakumaran, Reshma, et al. ‘Regulation of Mutant P53 Protein Expression’. *Frontiers in Oncology*, vol. 5, 2015, p. 284. *PubMed*, https://doi.org/10.3389/fonc.2015.00284.

Waldman, Alex D., et al. ‘A Guide to Cancer Immunotherapy: From T Cell Basic Science to Clinical Practice’. *Nature Reviews. Immunology*, vol. 20, no. 11, Nov. 2020, pp. 651–68. *PubMed*, https://doi.org/10.1038/s41577-020-0306-5.

Wang, Bin, et al. ‘Clinical Applications of STING Agonists in Cancer Immunotherapy: Current Progress and Future Prospects’. *Frontiers in Immunology*, vol. 15, Oct. 2024, p. 1485546. *DOI.org (Crossref)*, https://doi.org/10.3389/fimmu.2024.1485546.

Watson, Adam W., et al. ‘Breast Tumor Stiffness Instructs Bone Metastasis via Maintenance of Mechanical Conditioning’. *Cell Reports*, vol. 35, no. 13, June 2021, p. 109293. *DOI.org (Crossref)*, https://doi.org/10.1016/j.celrep.2021.109293.

Yang, Wan R., et al. ‘SQuIRE Reveals Locus-Specific Regulation of Interspersed Repeat Expression’. *Nucleic Acids Research*, vol. 47, no. 5, Mar. 2019, pp. e27–e27. *DOI.org (Crossref)*, https://doi.org/10.1093/nar/gky1301.

Yuan, Xiao, et al. ‘Single-Cell Profiling to Explore Immunological Heterogeneity of Tumor Microenvironment in Breast Cancer’. *Frontiers in Immunology*, vol. 12, 2021, p. 643692. *PubMed*, https://doi.org/10.3389/fimmu.2021.643692.

Zanconato, Francesca, et al. ‘Genome-Wide Association between YAP/TAZ/TEAD and AP-1 at Enhancers Drives Oncogenic Growth’. *Nature Cell Biology*, vol. 17, no. 9, Sept. 2015, pp. 1218–27. *DOI.org (Crossref)*, https://doi.org/10.1038/ncb3216.

Zhu, Gaoyang, et al. ‘Mutant P53 in Cancer Progression and Targeted Therapies’. *Frontiers in Oncology*, vol. 10, 2020, p. 595187. *PubMed*, https://doi.org/10.3389/fonc.2020.595187.

Zierhut, Christian, and Hironori Funabiki. ‘Regulation and Consequences of cGAS Activation by Self-DNA’. *Trends in Cell Biology*, vol. 30, no. 8, Aug. 2020, pp. 594–605. *PubMed*, https://doi.org/10.1016/j.tcb.2020.05.006.