Oncogenic TRIM37 in TNBC: From Molecular Mechanism to New Functions

Rachisan G. Djiake Tihagam

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Biochemistry and Molecular Genetics

University of Virginia

May 2025

Dedication

First and foremost, this thesis is dedicated to my family: my parents, Julienne Tchiablou, Jean-Jacques Ndam, Andre Tihagam; my sister Ulrich Leuffo; my brothers Junior Ndam and Valentin Ndam; my aunt Henriette Djieyap; and my nieces. This journey would not have been possible without your endless love, encouragement, and support. Thank you for always answering the calls, listening to my constant rants, and providing comfort. I love you all dearly.

To my dearest Donald Talla, I am profoundly grateful to have you in my life. Your love, patience, and belief in me have lifted me through the challenges and celebrated with me in moments of victory throughout this journey. I am forever grateful for your presence in my life.

Next, I would like to thank my Ph.D. advisor, Dr. Sanchita Bhatnagar, for her unwavering support, training, guidance, and expertise whenever I needed it. She has been instrumental in shaping my academic and personal growth. Throughout my graduate school journey, Dr. Bhatnagar has challenged me to think critically and work hard, and because of that, I am fully prepared for whatever comes next. Thank you again for that.

I would like to thank my thesis committee members, Drs. Jogender Singh, Jeffrey Smith, Ani Manichaikul, Roger Abounader, and Amy Bouton, for their amazing advice and feedback.

Dissertation Abstract

Oncogenic TRIM37 in TNBC- from molecular mechanism to new function

Rachisan G. Djiake Tihagam

Thesis advisor: Sanchita Bhatnagar, Ph.D.

Abstract

Triple negative breast cancer (TNBC) accounts for approximately 10–15% of all breast cancer diagnoses, and it carries the worst prognosis among all breast cancer subtypes. Clinically, the leading causes of death in TNBC patients are attributed to its high chemoresistance and the rapid spread of metastasis. These two factors are more prevalent in younger women of African descent, who experience disproportionately higher tumor burden compared to other races. Although tumorigenic drivers of TNBC are numerous and varied, the drivers of metastatic transition and chemoresistance remain largely unknown. Here, I characterized TRIM37 function in driving an aggressive TNBC biology, revealing its critical role in chemoresistance and metastatic progression. Interestingly, TNBC affects younger Black women (BW), who experience significantly higher incidence and lower survival rates compared to White women (WW). I demonstrated TRIM37 as a genetic driver of racial disparities in TNBC. My results show that the risk allele of rs57141087 in BW increases TRIM37 expression via NRF1-mediated interactions. Together, my results reveal that TRIM37 is a potential predictive biomarker and

therapeutic target for improving TNBC screening and treatment outcomes. My findings suggest that additional clinically relevant drivers of metastatic progression could exist, which could improve the likelihood of finding new treatments for TNBC patients. Therefore, future studies utilizing genome-wide genetic screens could be used toward this goal. Furthermore, new racially segregating variants associated with breast cancer risk could be identified using omics approaches. Based on my findings, future studies will focus on identifying new racially segregating genetic variants that could increase breast cancer risk.

Chapter 1: Triple-negative breast cancer (TNBC)

Chapter 2: Tripartite Motif (TRIM) Protein Family

Chapter 3: Oncogenic TRIM37 links chemoresistance and metastatic fate in triple-negative breast cancer

Chapter 4: The TRIM37 variant rs57141087 contributes to triple-negative breast cancer outcomes in Black women

Chapter 5: Conclusion

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
TNBC	Triple negative breast cancer
BW	Black Women
WW	White Women
TRIM37	Tripartite motif-containing protein 37
EMT	Epithelial-Mesenchymal Transition
NRF1	Nuclear respiratory factor 1
ER	Estrogen Receptor
PR	Progesterone Receptor
HER2	Human Epidermal Growth Factor Receptor 2
BL1	Basal-like subtype 1
BL2	Basal-like subtype 2
IM	Immunomodulatory
M	Mesenchymal
MSL	Mesenchymal Stem-Like
LAR	Luminal Androgen Receptor
PI3K	Phosphatidylinositol 3-kinase
CDK6	Cyclin-Dependent Kinase 6

FGFR1	Fibroblast growth factor receptor 1
Ki-67	Kiel 67
EGFR	Epidermal growth factor receptor
Th1/Th2	T helper cell type 1/T helper cell type 2
NK	Natural killer
DC	Dendritic Cell
IL-12	Interleukin-12
IL-6	Interleukin-6
TGF	Transforming Growth Factor
AR	Androgen receptors
BRCA1	Breast Cancer Gene 1
BRCA2	Breast Cancer Gene 2
NHEJ	Non-Homologous End Joining
TME	tumor microenvironment
PD-L1	Programmed Cell Death Ligand 1
CAF	Cancer-associated fibroblast
CAR	Chimeric antigen receptor
MEK	Mitogen-Activated Protein Kinase
MAPK	Mitogen-activated protein kinase
GWAS	Genome-wide association studies

SNP	Single nucleotide polymorphism
GLI1	Glioma-associated oncogene homolog 1
Notch1	Neurogenic locus notch homolog protein 1
PTEN	Phosphatase and tensin homolog
INPP4B	Inositol polyphosphate 4-phosphatase type II
PARP	Poly (ADP-ribose) polymerase
mTOR	Mammalian target of rapamycin
PTEN	Phosphatase and tensin homolog
INPP4B	Inositol polyphosphate 4-phosphatase
TRIM	Tripartite Motif
PHD	Plant Homeodomain
TM	Transmembrane
MATH	Meprin and TRAF Homology
HECT	Homologous to E6AP Carboxy Terminus
RING	Really Interesting New Gene
UTR	Untranslated region
TRAF	TNF receptor-associated factor
NLS	Nuclear localization signal
RNF2	RING finger protein 2
PRC1	Polycomb Repressive Complex 1

13

PRC2	Polycomb Repressive Complex 2
IGF	Insulin-like growth factor
ATM	Ataxia-telangiectasia mutated
E2F1	E2F Transcription Factor 1
STAT	Signal transducer and activator of transcription
EZH2	Enhancer of Zeste Homolog 2
GFI1	Growth factor independent-1
PAX5	Paired Box 5
EpCAM	Epithelial Cell Adhesion Molecule
CD24	Cluster of differentiation 24
CD44	Cluster of differentiation 44
Dox	Doxorubicin
TFBS	Transcription factor binding sites
MET	mesenchymal-epithelial transition
SMIs	Small molecule inhibitors

Chapter 1: Triple-negative breast cancer (TNBC)

In the United States, TNBC accounts for about 10-15% of all breast cancer cases, with a higher prevalence among younger women of African descent [1]. TNBC is more aggressive than other types of breast cancer, with a higher likelihood of rapid growth, metastasis to other organs, and a poorer prognosis, as well as a higher chance of recurrence [2]. Overall, the 5-year survival rate for TNBC patients is approximately 77%, which is significantly lower compared to 93% for other breast cancer subtypes in general [3]. The main challenges with TNBC include chemoresistance, higher metastasis rates, and lack of targeted therapies. These challenges contribute to the poor outcomes observed in TNBC patients.

1.1 TNBC Classification

TNBC tumors are characterized by the absence of estrogen (ERα), progesterone (PR) receptors, and human epidermal growth factor receptor 2 (HER2). Despite sharing these common characteristics, TNBC remains highly heterogeneous, with differences in molecular profiles, genetic mutations, and responses to treatment across individual tumors. A gene expression profiling and gene ontology analysis of 586 TNBC tumors [4] identified six distinct molecular subtypes of TNBC based on genomic, clinical, and pathologic features [4]. These subtypes include Basal-like subtype 1 (BL1) and subtype 2 (BL2), an immunomodulatory (IM) subtype, a

mesenchymal (M) subtype, a mesenchymal stem-like (MSL) subtype, and a luminal androgen receptor (LAR) subtype [4] (Fig. 1-1).

The BL1 subtype is characterized by increased expression of oncogenes like MYC, KRAS, PIK3CA, CDK6, AKT2, and FGFR1, as well as genes involved in cell proliferation, cell cycle, and DNA repair pathways [3-5]. The Ki67 levels are high in this subtype, indicative of an increased proliferation rate [6]. In contrast, the BL2 tumors have abnormal activation of signaling pathways, including EGFR, MET, NGF, Wnt/β-catenin, and IGF-1R pathways, often leading to a more aggressive clinical course and poorer outcomes [4].

The IM subtype tumors are enriched in genes associated with immune cell functions and signaling pathways, including the Th1/Th2, the NK cell, B cell receptor signaling, dendritic cell (DC), T cell receptor signaling, as well as the interleukin (IL)-12 and IL-6 pathways [1, 7]. Compared to other TNBC molecular subtypes, the IM subtype often presents a more favorable prognosis due to its active immune microenvironment [8].

The M subtype exhibits abnormal activation of pathways associated with epithelial-mesenchymal transition (EMT), including those involved in cell migration, extracellular matrix–receptor interactions, and differentiation processes, such as the Wnt pathway, anaplastic lymphoma kinase pathway, and transforming growth factor (TGF)-β signaling [1, 7]. EMT is a process that allows cancer cells

to have invasive and metastatic properties. Therefore, TNBC patients with M tumors are typically associated with more aggressive disease, poorer prognosis, and a higher rate of recurrence.

MSL subtypes share similarities with the M subtype in that these tumors also exhibit mesenchymal characteristics and a tendency toward aggressiveness [9]. However, while the M subtype displays overexpression of proliferation genes, the MSL subtype exhibits low levels of these genes and high expression of stemness-related genes, mesenchymal stem cell-specific markers, and genes involved in angiogenesis and growth factor pathways [7].

Finally, the LAR subtype features a distinct gene expression profile that differentiates it from other TNBC subtypes [7]. It is defined by high expression of both the androgen receptor (AR) and its downstream target genes, suggesting a reliance on androgen signaling pathways. As a result, this subtype may respond to androgen receptor-targeted therapies, which can influence steroid metabolism and synthesis [7]. These TNBC subtypes underscore the complexity and heterogeneity of TNBC, highlighting the need for personalized treatment strategies that target specific molecular features of the tumors.

In their classification of TNBC cases, Lehmann et al. identified several subtypes with distinct gene expression profiles, with the following distribution: 18% BL1, 11% BL2, 21% IM, 21% M, 8% MSL, 9% LAR, and 12% unclassified

[10] (Fig.1-1). Masuda et al. conducted a prognostic analysis of various TNBC

subtypes and discovered that the LAR subtype had a better distant metastasis-free

survival rate and overall survival (OS). In contrast, the M and BL2 subtypes

exhibited poorer outcomes. Specifically, the 3-year recurrence rates for the M and

BL2 subtypes were significantly higher than the LAR subtype [7, 11].
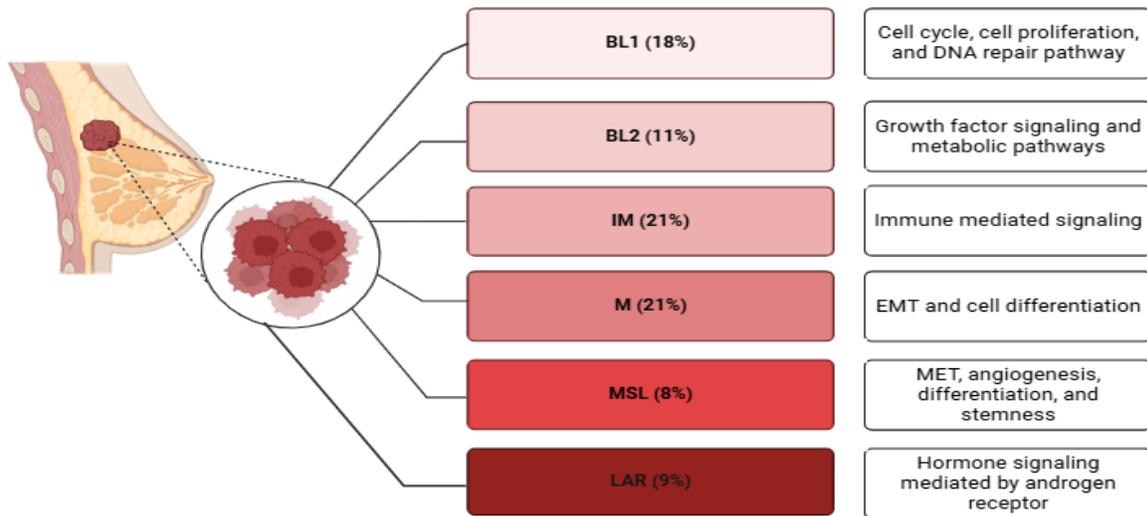
Figure 1-1. Molecular classification of TNBC and their characteristics. The schematic shows six distinct TNBC subtypes, each associated with distinct gene expression profiles and clinical features. These subtypes include: BL1 (basal-like 1), BL2 (basal-like 2), IM (immunomodulatory), M (mesenchymal), MSL (mesenchymal stem-like), and LAR (luminal androgen receptor). (Adapted from Lehmann et al., 2016)

1.2 Genetic Insights into TNBC

Genetic insights into TNBC have revealed the complexity of this aggressive

disease and its resistance to treatment. Some key genetic markers that influence

TNBC prognosis and help predict the most effective therapies include mutations in

BRCA1/2, TP53, PIK3CA, and the loss of PTEN/INPP4B (Fig. 1-2) [5, 12, 13].

Mutations in BRCA1/2 are highly prevalent in TNBC patients and include deletions, insertions, and single-nucleotide substitutions in both coding and non-coding sequences. BRCA1/2 mutations influence cancer biology by impairing the DNA repair mechanisms, specifically through the homologous recombination repair pathway (HR). These mutations lead to an inability to properly repair DNA damage, resulting in genomic instability [14]. As a result, cancers with BRCA1/2 mutations tend to accumulate more genetic alterations, which contribute to the aggressive nature of the disease. However, this same defect in DNA repair also presents therapeutic opportunities. For instance, patients with BRCA1/2 mutations often show increased sensitivity to chemotherapeutic drugs, such as platinum-based treatments like cisplatin, which induce DNA damage [14]. These drugs work by causing DNA damage in cancer cells, ultimately leading to cell death.

Moreover, PARP inhibitors, such as olaparib and talazoparib, have shown significant efficacy in treating BRCA-mutated TNBC by exploiting the cancer's impaired DNA repair mechanisms [15]. PARP inhibitors are a class of targeted therapies designed to interfere with the PARP enzyme, which plays a crucial role in repairing single-strand breaks in DNA. By inhibiting PARP, these drugs prevent the repair of additional DNA damage, accumulating unrepaired breaks and ultimately causing cancer cell death [15]. TP53 mutations are found in about 80% of all TNBC cases, which contributes to the aggressive nature of TNBC, increasing the

likelihood of metastasis and leading to a worse overall prognosis compared to other breast cancer subtypes [5]. Moreover, TP53 mutations are associated with resistance to certain therapies, including standard chemotherapy, making managing TNBC particularly challenging.

PIK3CA mutations are found in approximately 10% of TNBC patients [16], which activate the PI3K/AKT/mTOR signaling pathway to accelerate tumor growth and promote resistance to therapy. Approximately 30-50% of TNBC patients have mutations or loss of expression in either the phosphatase and tensin homolog (PTEN) or inositol polyphosphate 4-phosphatase (INPP4B) gene [17]. PTEN and INPP4B negatively regulate PI3K signaling, and their loss of function leads to unchecked activation of the PI3K/AKT/mTOR signaling pathway. This unchecked activation further accelerates tumor cell proliferation, resulting in a poor prognosis [17].
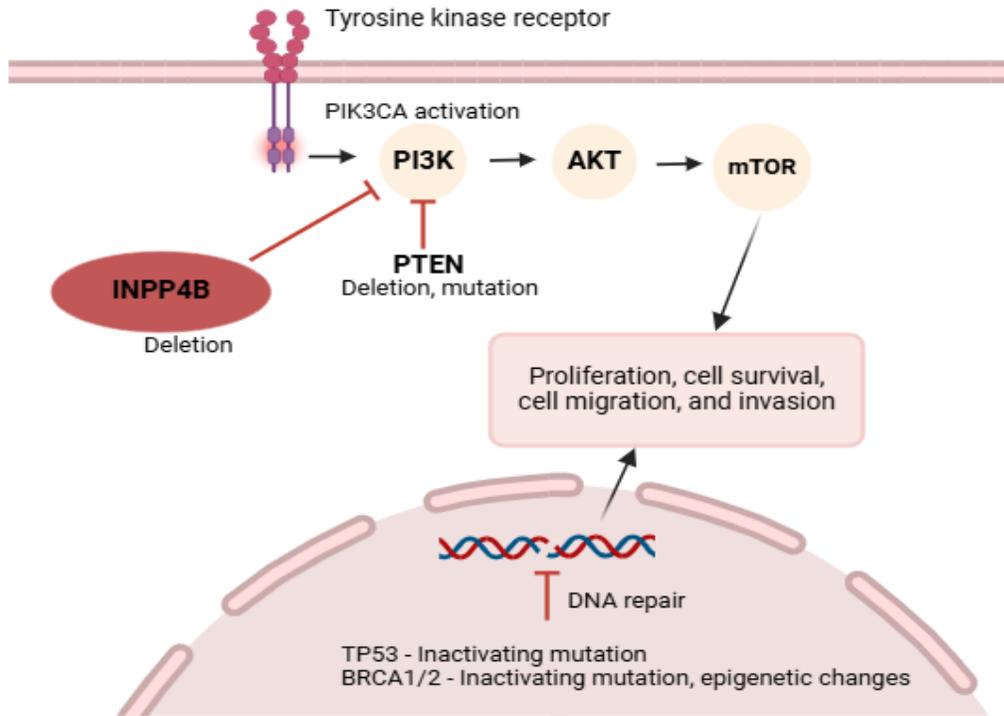
Figure 1-2. Genetic Markers and Their Influence on TNBC Prognosis And/Or Prediction. This figure illustrates key genetic alterations that include mutations in BRCA1/2, TP53, and the loss of PTEN/INPP4B, all of which play a critical role in determining clinical outcomes and response to treatment in TNBC. Inactivating mutations in BRCA1/2 impair DNA repair mechanisms, leading to genomic instability and increased sensitivity to DNA-damaging agents. TP53-inactivating mutations disrupt cell cycle regulation and are associated with poor prognosis and resistance to certain chemotherapies. Additionally, the loss of PTEN/INPP4B results in deregulating the PI3K/AKT pathway, driving tumor progression and complicating treatment response. (Adapted from Sporikova et al., 2018).

1.3 Role of Epigenetic Alterations in TNBC

While genetic alterations, such as mutations, amplifications, or deletions, drive the TNBC initiation and progression, they are not the sole contributors to the disease's aggressive behavior. Epigenetic alterations, which regulate gene expression without altering the DNA sequence, also play a pivotal role in TNBC. These epigenetic changes, including histone modifications, microRNA (miRNAs)

expression, and long non-coding RNAs (lncRNAs), can influence tumorigenesis by altering gene expression and chromatin structure [18].

Histone modifications, including acetylation, methylation, and ubiquitination, involve the addition or removal of specific functional groups, such as acetyl groups (-COCH3), methyl groups (-CH3), or ubiquitin molecules, on histone proteins. These modifications can alter the chromatin structure and influence gene expression. Depending on the type of modification and its location on the histone, these changes can either activate or repress gene transcription. Histone modifications are crucial for regulating genes involved in cell proliferation, survival, and metastasis, contributing to TNBC's aggressive behavior [18].

In addition to histone modifications, gene expression is also regulated by miRNAs and lncRNAs, which modulate gene expression post-transcriptionally. miRNAs are small, non-coding RNA molecules that bind to the 3' untranslated regions (UTRs) of target messenger RNAs (mRNAs), resulting in their degradation or inhibition of translation [18]. lncRNAs also play a crucial role in regulating gene expression in TNBC by interacting with chromatin [18]. These regulatory mechanisms can result in the overexpression of oncogenes or the downregulation of tumor suppressor genes, both of which contribute to tumorigenesis and metastasis in TNBC [18]. Together, these epigenetic changes contribute to the molecular heterogeneity of TNBC, making it a highly complex and difficult-to-treat subtype

of breast cancer. In many cases, epigenetic modifications work with in concert with genetic alterations, either enhancing the effects of genetic mutations or compensating for the loss of function in tumor suppressor genes. The intricate interplay between genetic and epigenetic alterations highlights the complexity of TNBC and underscores the need for a better approach to understanding its molecular basis and developing effective treatments.

1.4 Risk factors in TNBC

In addition to genetic and epigenetic factors, several other elements contribute to the development of TNBC. Key risk factors for TNBC include age, lifestyle choices, and race.

a. Age:

TNBC is more commonly diagnosed in women under the age of 50, with a higher prevalence among premenopausal women [21]. Younger women tend to develop more aggressive forms of TNBC than older women, which can make it more challenging to treat and manage, resulting in poorer survival rates.

b. Lifestyle choices:

Lifestyle factors are everyday personal habits and choices that can be controlled and directly impact an individual's health and well-being. Lifestyle factors that have been linked to an increased risk of developing TNBC include

being overweight or obese, lack of physical activity, smoking, excessive alcohol consumption, chronic stress, and poor sleeping patterns [22].

c. Race:

Studies have shown that TNBC incidence and mortality rates vary significantly across racial and ethnic groups. TNBC occurs at a disproportionately higher rate in BW compared to WW, and the clinical outcomes for these women tend to be much more severe. BW is twice as likely to be diagnosed with TNBC as WW. Furthermore, they tend to develop TNBC at younger ages and often present with more advanced stages of the disease at diagnosis [23]. In addition to higher incidence rates, BW also experiences worse outcomes in terms of survival. The 5-year relative survival rate for BW with TNBC is approximately 14%, compared to 36% for other races [24]. The racial disparities in TNBC outcomes are attributed to a complex interplay of socio-economic and biological factors.

1.5 Socio-economic factors

Socio-economic factors refer to a combination of social, economic, and environmental conditions in which individuals live, which can affect their overall access to resources, healthcare, and opportunities for healthy living. Socioeconomic factors, such as access to healthcare, income, and education, significantly contribute to TNBC disparity [25]. In the United States (U.S.), between 2014 and 2016, the

average household income in the U.S. was $ 70,000 compared to $48,000 for Black households [26]. BW experience higher levels of poverty, lower educational attainment, higher unemployment rates, and are more likely to reside in economically disadvantaged neighborhoods [26].

a. Access to healthcare:

Individuals from lower socio-economic backgrounds, such as BW, often have limited access to health insurance, preventative care, and timely medical treatment, which can delay the diagnosis of TNBC [26]. Additionally, late-stage diagnoses are common in underserved communities, where access to mammograms and early TNBC screening may be limited [26]. Because early detection of TNBC is crucial for improving survival rates, delayed or missed screenings tend to result in more aggressive disease and poor survival outcome.

b. Income and Education:

Low-income individuals often face financial barriers that prevent them from accessing quality healthcare, including preventive services like screenings and early detection. These individuals may also have limited access to healthy food or resources needed to maintain a healthy lifestyle, such as gym memberships or wellness programs [25]. As a result, they may be more likely to engage in behaviors that increase the risk of developing TNBC, such as poor dietary choices, smoking, and physical inactivity [25, 26]. Additionally, financial barriers may limit their

11

ability to seek timely medical attention or follow through with recommended treatments, increasing the risks. Moreover, lower levels of education can lead to a lack of awareness about cancer prevention and risk factors, which may contribute to delays in diagnosis. In contrast, individuals with higher education levels and greater financial stability typically have better access to healthcare, a deeper understanding of risk factors, and healthier habits, all of which can reduce their risk of developing TNBC [25, 26].

1.6 Biological factors

While socio-economic factors play a significant role in TNBC disparities, studies also indicate that biological differences, such as tumor genetics and gene expression patterns linked to race, also contribute to disparities in TNC prevalence, aggressiveness, and treatment outcomes.

Biological differences between racial groups also affect TNBC susceptibility, disease progression, and response to treatment. Studies have shown that genetic mutations, particularly in BRCA1 and BRCA2, are more prevalent in BW and are strongly associated with an increased risk of TNBC [25]. Additionally, TP53 mutations, common in TNBC, are more frequently observed in BW and contribute to poorer prognosis due to their role in promoting genomic instability and resistance to chemotherapy [27]. A genome-wide association studies (GWAS) study has discovered single nucleotide polymorphisms (SNPs), such as rs66664032,

rs10069690, and rs12964508, that are more dominant in BW and correlate with increased TNBC risk [28]. Epigenetic alterations, including the overexpression of pro-inflammatory cytokines like IL-6, have been linked to worse clinical outcomes and metastasis in BW with TNBC [25]. Furthermore, glioma-associated oncogene homolog 1(GLI1), a transcription factor involved in the Hedgehog signaling pathway, and neurogenic locus notch homolog protein 1 (Notch1), a key regulator of cell fate and differentiation, are more strongly activated in TNBC tumors from BW, contributing to the disparity in TNBC outcomes [24]. These biological factors, coupled with socio-economic and healthcare-related challenges, contribute to the disproportionate burden of TNBC in BW, ultimately leading to poorer survival rates and highlighting the urgent need for targeted interventions and equitable healthcare access.

1.7 Current treatment options and challenges for TNBC

The current treatment approach for TNBC is multifaceted and often involves a combination of surgery, radiation, chemotherapy, targeted therapy, immunotherapy, and, in recent years, antibody-drug conjugates (ADCs) (Fig 1-3) [29].

Surgery is often the first line of treatment for TNBC patients, particularly in cases diagnosed at an early stage when the cancer is localized. The primary goal is to remove the tumor, affected breast tissues, and lymph nodes from the body to

minimize the risk of cancer recurrence. The surgical options for TNBC depend on the tumor's size, location, and stage. A lumpectomy, which removes the tumor and some surrounding tissue, is typically considered for smaller tumors and early-stage breast cancer [30]. This treatment option is typically considered for smaller tumors and early-stage breast cancer. In cases where the tumor is large or the cancer has spread to multiple areas of the breast, a mastectomy, which involves the complete removal of one or both breasts, is preferred [30]. In addition, lymph node assessment is critical, and procedures such as sentinel lymph node biopsy or axillary lymph node dissection are often performed to evaluate whether the cancer has spread to lymph nodes. The timing and type of surgery are influenced by the tumor's response to preoperative treatment and the need for further interventions [30, 31].

Radiation therapy is typically recommended after surgical intervention for TNBC patients, particularly those who have undergone breast-conserving surgery [32]. The primary aim of radiation is to eliminate any remaining cancer cells in the breast or surrounding tissues, thereby reducing the risk of local recurrence. This treatment is especially important for patients with larger tumors or those with lymph node involvement. Although radiation therapy is generally well-tolerated, potential side effects can include skin irritation, fatigue, and changes in breast appearance. Nonetheless, the benefits of radiation therapy in improving overall

survival and reducing recurrence rates make it an integral part of the treatment plan for patients with TNBC, particularly following surgical intervention [33].

Chemotherapy is the primary therapeutic option for TNBC patients. It is commonly administered in two settings: neoadjuvant (before surgery) and adjuvant (after surgery). In the neoadjuvant setting, chemotherapy is given to shrink tumors, making them more manageable for surgical removal and potentially allowing for breast-conserving surgery. Standard regimens typically involve combinations of doxorubicin, taxanes like paclitaxel, and carboplatin [34]. In the adjuvant setting, chemotherapy aims to eliminate residual cancer cells and reduce the risk of recurrence following surgery.

Targeted therapy represents an evolving treatment option for TNBC patients. Unlike conventional chemotherapy, which affects both cancerous and healthy cells, targeted therapies are designed to specifically attack the molecular targets present within the tumor, such as genetic mutations or altered signaling pathways, making them particularly effective in treating solid tumors like TNBC. Certain targeted therapies have shown promise in clinical settings. PARP inhibitors, such as olaparib and talazoparib, are designed for patients with BRCA1/2 mutations, capitalizing on the tumors' impaired DNA repair mechanisms to induce cancer cell death [15]. Mitogen-activated protein Kinase (MEK) has also been explored as a potential targeted therapy option for treating TNBC patients. MEK is a key enzyme in the

MAPK signaling pathway, vital for cell proliferation, differentiation, and survival. By inhibiting MEK, these drugs aim to block signals that promote cancer cell growth, making them a potential therapeutic option for TNBC [35]. PI3K inhibitors are also being explored for their potential effectiveness in TNBC characterized by PIK3CA mutations [35].

In recent years, immunotherapy has become a leading treatment option for many cancers. TNBC tumors are characterized by a higher presence of tumor-infiltrating lymphocytes and elevated expression of PD-L1 protein, especially when compared to other subtypes of breast cancer[8]. For patients with metastatic TNBC, immunotherapy combined with chemotherapy has demonstrated improved survival outcomes compared to chemotherapy alone. Immune checkpoint inhibitors like pembrolizumab and atezolizumab are now used with chemotherapy to boost the immune system's ability to attack only cancer cells selectively. This approach has shown significant efficacy, particularly in patients with high PD-L1 expression, which can serve as a biomarker for response to treatment. Atezolizumab (Tecentriq), a PD-L1 monoclonal antibody, has long-term benefits for approximately 10% of TNBC patients [36]. Enhancing the responsiveness to anti-PD-1/PD-L1 therapies has the potential to lower mortality rates among patients with triple-negative breast cancer (TNBC), providing greater reassurance and hope for those affected by this aggressive cancer subtype. In addition to checkpoint

inhibitors, chimeric antigen receptor (CAR) T-cell therapy is another promising immunotherapy. This treatment aims to activate the patient's immune system to attack TNBC cells that express specific antigens not found in healthy cells [37]. While targeted therapies promise more personalized treatment options, they are generally considered after standard chemotherapy regimens and are still being evaluated in clinical trials to determine their optimal use.

In recent years, the Food and Drug Administration's (FDA) approval of antibody-drug conjugates (ADCs) has marked a significant advancement in the treatment of TNBC, offering new hope for patients with this aggressive breast subtype. ADCs are an emerging class of targeted therapies consisting of a recombinant monoclonal antibody that targets a specific cell surface antigen, conjugated to a cytotoxic agent through either a cleavable or non-cleavable synthetic linker [38]. The antibody targets a specific protein on the surface of cancer cells, delivering the chemotherapy drug directly to the tumor, which minimizes damage to surrounding healthy tissues and enhances the effectiveness of the treatment. ADCs address the limitations of traditional chemotherapy, including the narrow therapeutic window and the development of resistance mechanisms [38]. In metastatic TNBC, the two FDA-approved ADCs are Sacituzumab govitecan (Trodelvy) and Trastuzumab deruxtecan (Enhertu) [39]. Trodelvy is an ADC that targets trophoblast cell surface antigen 2 (TROP2), a protein frequently

17

overexpressed on the surface of TNBC cells. By binding to TROP2, Trodelvy delivers the chemotherapeutic agent SN-38, a potent antineoplastic drug, directly to the cancer cells, thereby reducing the risk of systemic toxicity [40]. Enhertu, initially approved for HER2-positive breast cancer, has shown efficacy in TNBC patients with HER2-low expression [41]. Trastuzumab binds to HER2 on the surface of cancer cells and delivers the chemotherapeutic agent Deruxtecan directly to the cancer cells, damaging their DNA and killing them [41]. The approval of these ADCs represents a significant step forward in the fight against TNBC by providing more direct, targeted, effective therapies that can improve outcomes, particularly in patients with metastatic disease or those with limited treatment options. However, as with all therapies, these ADCs do not work for every patient. Therefore, ongoing research must be done to develop more ADCs and identify which patients are most likely to benefit from these advanced treatments.

Despite recent advancements in the treatment of TNBC, new drugs are urgently needed for several reasons. First, TNBC is known for its aggressiveness, characterized by rapid growth, a high likelihood of recurrence, and resistance to many traditional therapies. The lack of targeted receptors for estrogen, progesterone, and HER2 makes it challenging to treat, and many patients relapse after initial treatment. Additionally, the limited response to current therapies, such as chemotherapy, surgery, radiation, and ADCs, is a major challenge. Many TNBC

patients do not respond well to these treatments, and the side effects tend to be severe, further complicating treatment, especially with the increased risk of metastasis, particularly to the lungs, liver, and brain.

Furthermore, the heterogeneity within TNBC adds complexity, as therapies effective for one subtype may not be effective for another, emphasizing the need for a more personalized approach. Notably, not all TNBC patients respond to these treatments, underscoring the need for more effective combination therapies or novel agents to overcome resistance mechanisms. In conclusion, surgery, radiation, chemotherapy, targeted therapy and immunotherapy have significantly improved TNBC treatment outcomes. However, ongoing research and new drug development remain essential to address the challenges posed by this aggressive breast cancer subtype. Continued efforts to understand the different molecular mechanisms of TNBC and identify new therapeutic targets will be crucial for improving outcomes for patients with TNBC.
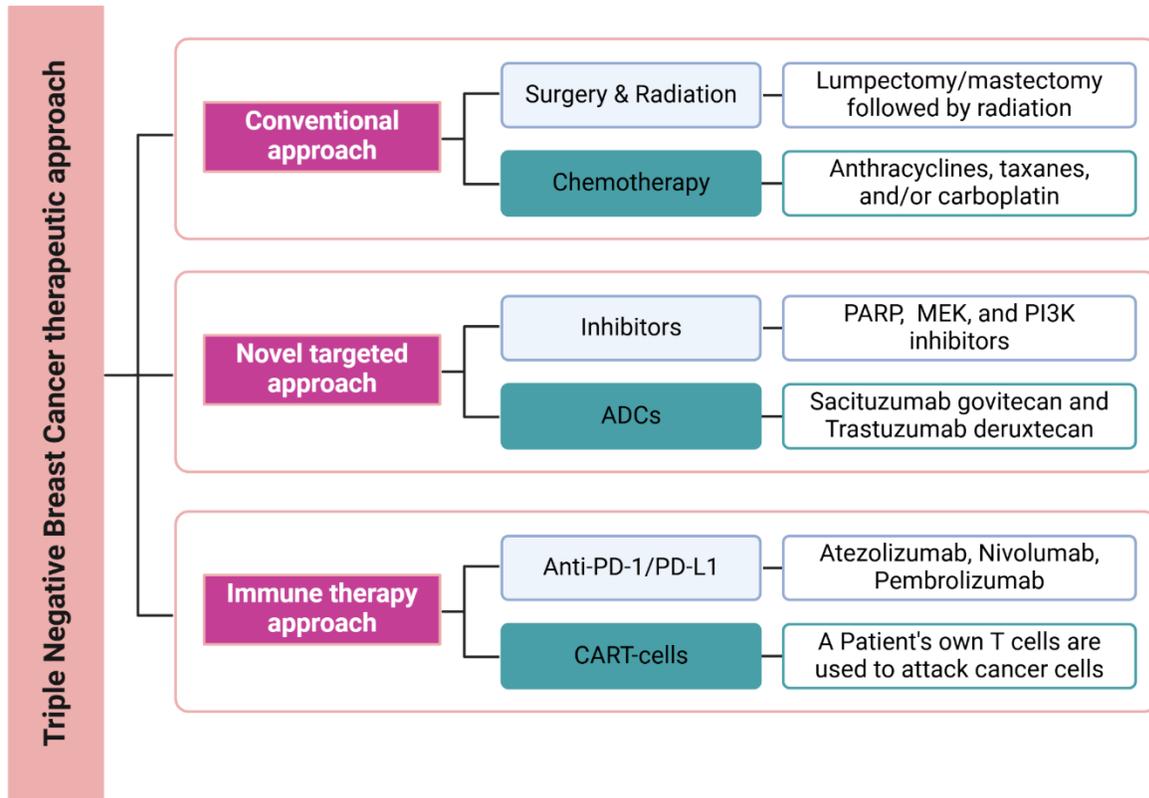
Figure 1-3. Various therapeutic strategies for treating TNBC. This figure illustrates the different approaches used in treating TNBC, including the conventional approach, novel targeted therapies, and immunotherapy. The conventional approach typically involves surgery, radiation, and chemotherapy. The novel targeted approach focuses on therapies designed to target specific molecular pathways and cell surface receptors involved in TNBC progression. Immune therapy stimulates the patient's immune system to recognize better and attack TNBC cells. (Adapted from Xiong et al., 2024)

1.8 References

1.   Zagami, P. and L.A. Carey, Triple negative breast cancer: Pitfalls and progress. NPJ Breast Cancer, 2022. 8(1): p. 95.

2.   Newman, L. and E. Mitchell, Disparities in triple negative breast cancer. J Natl Med Assoc, 2023. 115(2S): p. S8-S12.

3.   Hsu, J.Y., C.J. Chang, and J.S. Cheng, Survival, treatment regimens and medical costs of women newly diagnosed with metastatic triple-negative breast cancer. Sci Rep, 2022. 12(1): p. 729.

4.   Lehmann, B.D., et al., Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest, 2011. 121(7): p. 2750-67.

5.   Derakhshan, F. and J.S. Reis-Filho, Pathogenesis of Triple-Negative Breast Cancer. Annu Rev Pathol, 2022. 17: p. 181-204.

6.   Healey, M.A., et al., Assessment of Ki67 expression for breast cancer subtype classification and prognosis in the Nurses' Health Study. Breast Cancer Res Treat, 2017. 166(2): p. 613-622.

7.   Yin, L., et al., Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res, 2020. 22(1): p. 61.

8.   Liu, Y., et al., Correction: Advances in immunotherapy for triple-negative breast cancer. Mol Cancer, 2023. 22(1): p. 154.

9.   Dai, X., et al., Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res, 2015. 5(10): p. 2929-43.

10.   Lehmann, B.D., et al., Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One, 2016. 11(6): p. e0157368.

11.   Masuda, H., et al., Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin Cancer Res, 2013. 19(19): p. 5533-40.

12.   Mallick, S., A.K. Duttaroy, and S. Dutta, The PIK3CA gene and its pivotal role in tumor tropism of triple-negative breast cancer. Transl Oncol, 2024. 50: p. 102140.

13.   Stevens, K.N., C.M. Vachon, and F.J. Couch, Genetic susceptibility to triple-negative breast cancer. Cancer Res, 2013. 73(7): p. 2025-30.

14.   Chai, Y., et al., Homologous Recombination Deficiency (HRD) and BRCA 1/2 Gene Mutation for Predicting the Effect of Platinum-Based Neoadjuvant Chemotherapy of Early-Stage Triple-Negative Breast Cancer (TNBC): A Systematic Review and Meta-Analysis. J Pers Med, 2022. 12(2).

15.   Barchiesi, G., et al., Emerging Role of PARP Inhibitors in Metastatic Triple Negative Breast Cancer. Current Scenario and Future Perspectives. Front Oncol, 2021. 11: p. 769280.

16.   Sporikova, Z., et al., Genetic Markers in Triple-Negative Breast Cancer. Clin Breast Cancer, 2018. 18(5): p. e841-e850.

17.   Reed, D.E. and K.M. Shokat, INPP4B and PTEN Loss Leads to PI-3,4-P2 Accumulation and Inhibition of PI3K in TNBC. Mol Cancer Res, 2017. 15(6): p. 765-775.

18.   Yang, J., et al., Epigenetic regulation in the tumor microenvironment: molecular mechanisms and therapeutic targets. Signal Transduct Target Ther, 2023. 8(1): p. 210.

19.   Yoshida, K. and Y. Miki, Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci, 2004. 95(11): p. 866-71.

20.   Jain, A., A. Barge, and C.N. Parris, Combination strategies with PARP inhibitors in BRCA-mutated triple-negative breast cancer: overcoming resistance mechanisms. Oncogene, 2024.

21.   Dolle, J.M., et al., Risk factors for triple-negative breast cancer in women under the age of 45 years. Cancer Epidemiol Biomarkers Prev, 2009. 18(4): p. 1157-66.

22.   Kumar, N., et al., The unique risk factor profile of triple-negative breast cancer: a comprehensive meta-analysis. J Natl Cancer Inst, 2024. 116(8): p. 1210-1219.

23.   Adrada, B.E., et al., Triple-Negative Breast Cancer: Histopathologic Features, Genomics, and Treatment. Radiographics, 2023. 43(10): p. e230034.

24.   Siddharth, S., et al., Concomitant activation of GLI1 and Notch1 contributes to racial disparity of human triple negative breast cancer progression. Elife, 2021. 10.

25.   Prakash, O., et al., Racial Disparities in Triple Negative Breast Cancer: A Review of the Role of Biologic and Non-biologic Factors. Front Public Health, 2020. 8: p. 576964.

26.   Ubbaonu, C.D., et al., Disparities in Receipt of National Comprehensive Cancer Network Guideline-Adherent Care and Outcomes among Women with Triple-Negative Breast Cancer by Race/Ethnicity, Socioeconomic Status, and Insurance Type. Cancers (Basel), 2023. 15(23).

27.   Huszno, J. and E. Grzybowska, TP53 mutations and SNPs as prognostic and predictive factors in patients with breast cancer. Oncol Lett, 2018. 16(1): p. 34-40.

28.   Jia, G., et al., Genome-wide association analyses of breast cancer in women of African ancestry identify new susceptibility loci and improve risk prediction. Nat Genet, 2024. 56(5): p. 819-826.

29.   Xiong, N., H. Wu, and Z. Yu, Advancements and challenges in triple-negative breast cancer: a comprehensive review of therapeutic and diagnostic strategies. Front Oncol, 2024. 14: p. 1405491.

30.   Obidiro, O., G. Battogtokh, and E.O. Akala, Triple Negative Breast Cancer Treatment Options and Limitations: Future Outlook. Pharmaceutics, 2023. 15(7).

31.   Cipolla, C., et al., Comprehensive Axillary Management of Clinically Node-Positive (cN+) Breast Cancer Patients: A Narrative Review on Neoadjuvant Chemotherapy. Cancers (Basel), 2024. 16(19).

32.   Yao, Y., et al., Radiotherapy after surgery has significant survival benefits for patients with triple-negative breast cancer. Cancer Med, 2019. 8(2): p. 554-563.

33.   Steward, L.T., et al., Impact of radiation therapy on survival in patients with triple-negative breast cancer. Oncol Lett, 2014. 7(2): p. 548-552.

34.   Mandapati, A. and K.E. Lukong, Triple negative breast cancer: approved treatment options and their mechanisms of action. J Cancer Res Clin Oncol, 2023. 149(7): p. 3701-3719.

35.   Yang, R., et al., Therapeutic progress and challenges for triple negative breast cancer: targeted therapy and immunotherapy. Mol Biomed, 2022. 3(1): p. 8.

36.   Li, L., et al., Immunotherapy for Triple-Negative Breast Cancer: Combination Strategies to Improve Outcome. Cancers (Basel), 2023. 15(1).

37.   Corti, C., et al., CAR-T cell therapy for triple-negative breast cancer and other solid tumors: preclinical and clinical progress. Expert Opin Investig Drugs, 2022. 31(6): p. 593-605.

38.   Dri, A., et al., Breaking barriers in triple negative breast cancer (TNBC) -

Unleashing the power of antibody-drug conjugates (ADCs). Cancer Treat Rev,

2024. 123: p. 102672.

39.   Keskinkilic, M. and R. Sacks, Antibody-Drug Conjugates in Triple Negative

Breast Cancer. Clin Breast Cancer, 2024. 24(3): p. 163-174.

40.   Santi, D.V., L. Cabel, and F.C. Bidard, Does sacituzumab-govitecan act as a

conventional antibody drug conjugate (ADC), a prodrug of SN-38 or both? Ann

Transl Med, 2021. 9(14): p. 1113.

41.   Bardia, A. and G. Viale, HER2-Low Breast Cancer-Diagnostic Challenges

and Opportunities for Insights from Ongoing Studies: A Podcast. Target Oncol,

2023. 18(3): p. 313-319.

Chapter 2: Tripartite Motif (TRIM) Protein Family

2.1 Overview of the TRIM Family Proteins

The Tripartite Motif (TRIM) family consists of a large and diverse group of multidomain proteins characterized by the presence of three highly conserved amino-terminal domains: A RING finger domain (R), B-box domain 1 and/or B-box domain 2 (B1 and/or B2), and a coiled-coil region (CC) [1]. The carboxyl terminus domain of TRIM proteins is highly variable, which facilitates the classification of these proteins into distinct subfamilies and dictates the structure and function of each subfamily [1, 2]. TRIM proteins are classified into several subfamilies, including the PHD-BROMO, PRY/SPRY, NHL, TM, and MATH subfamilies, each of which has unique structural and functional characteristics (Fig.2-1A-B) [3, 4]. These are just a few of the diverse subfamilies within the TRIM family.

2.2 Classification of TRIM Protein Subfamilies

The PHD-BROMO subfamily includes proteins that contain a PHD finger (Plant Homeodomain) and a BROMO domain [5]. The PHD finger is typically located at the carboxyl terminus and plays a crucial role in chromatin recognition and transcriptional regulation by interacting with histone modifications [1, 6]. The BROMO domain, often adjacent to the PHD finger, is involved in protein-protein interactions and binds acetylated lysine residues on histones, further influencing

gene expression [7]. Together, the PHD and BROMO domains are crucial for regulating transcription and chromatin remodeling.

Proteins in the PRY/SPRY subfamily contain the PRY/SPRY domain at the carboxyl terminus. This domain has a characteristic β-sheet structure that mediates protein-protein interactions. It is frequently involved in immune responses, particularly in antiviral defense [4]. Additionally, the amino-terminal tripartite domains act as scaffolds for further functional domains, increasing the specificity of PRY/SPRY proteins in immune surveillance.

The NHL (NCL-1, HT2A, and Lin-41) subfamily proteins contain an NHL domain at the carboxyl terminus, which is involved in RNA binding and regulation [8]. The NHL domain is a beta-propeller structure that mediates interactions with RNA or other cellular targets [9]. This subfamily of proteins plays key roles in RNA processing and splicing and regulates the stability and translation of mRNAs, allowing the NHL subfamily to impact cellular gene regulation [8, 9].

The Transmembrane (TM) subfamily is characterized by a transmembrane domain at the carboxyl terminus, which enables these proteins to associate with cellular membranes [10]. The transmembrane domain is crucial for mediating interactions between TRIM and membrane-bound proteins or lipids, often regulating membrane trafficking and cell signaling pathways [10]. The amino-terminal remains consistent, while the carboxyl-terminal transmembrane region

allows the TM subfamily to specialize in signaling pathways associated with membrane dynamics and protein localization to cellular membranes.

The Meprin and TRAF Homology (MATH) domain at the carboxyl terminus defines the MATH subfamily of TRIM proteins. Proteins in this subfamily mediate protein-protein interactions, particularly in pathways related to cell growth, differentiation, and apoptosis [11]. The MATH domain often mediates signaling cascades by interacting with other signaling proteins, such as kinases or scaffold proteins [12]. Additionally, the MATH subfamily also plays a critical role in regulating the NF-κB pathway, which is essential for immune response and inflammation [13].
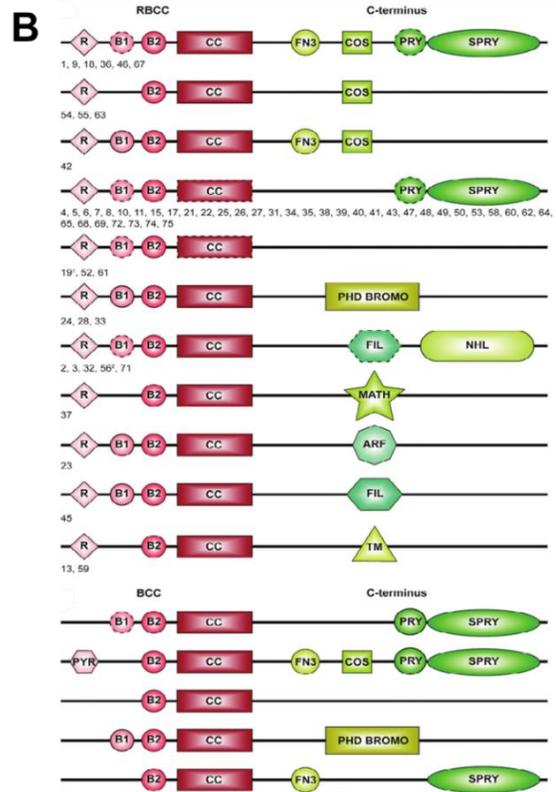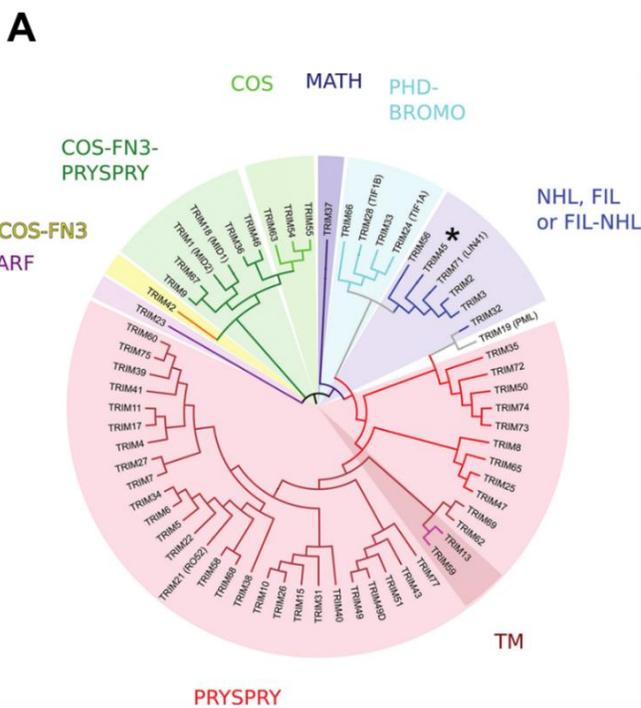
Figure 2-1. Phylogeny of human TRIM proteins. (A) The phylogenetic tree based on the tripartite motif alone mostly reflects the natural classification of TRIM proteins based on the C-terminal domains. (B) The RBCC and BCC motif-containing TRIM proteins are shown, with numbers corresponding to individual TRIM genes. (Copied from Williams et al., 2019).

2.3 TRIM Proteins as E3 Ubiquitin Ligases

As of today, at least sixty-six TRIM proteins have been identified in humans [4]. These proteins are classified into subfamilies, each characterized by distinct structural and functional features that enable them to perform specialized roles in processes such as chromatin remodeling, immune regulation, and cell signaling. Despite these differences, all TRIM proteins share a common function: they act as E3 ubiquitin ligases [4]. This activity is mediated by the RING domain, located in the amino-terminal region of TRIM proteins, which facilitates the transfer of ubiquitin to target proteins. The ubiquitination process marks these proteins for degradation by the proteasome or directs them to other critical cellular processes, such as protein trafficking, DNA repair, and signal transduction [2, 14, 15, 16]. By regulating protein turnover, TRIM proteins play a key role in maintaining cellular homeostasis and influencing biological processes such as cell cycle progression, apoptosis, differentiation, and immune responses [17].

Ubiquitination is a three-step process involving the sequential action of three key enzymes. First, the ubiquitin-activating enzyme (E1) activates ubiquitin in an ATP-dependent manner, forming a thioester bond between the C-terminus of ubiquitin and a catalytic cysteine residue within the E1 enzyme [21]. Next, the

activated ubiquitin is transferred to the ubiquitin-conjugating enzyme (E2), where it is linked to the E2's catalytic cysteine, effectively shuttling the ubiquitin for further modification [21]. Finally, the E3 ligase, such as TRIM proteins, facilitates the transfer of ubiquitin from E2 to a lysine residue on the target protein. This process can lead to the formation of polyubiquitin chains, influencing the substrate's fate [21] (Fig. 2-2).

E3 ligases can be divided into two major classes based on their catalytic domains: HECT (Homologous to E6AP Carboxy Terminus) and RING (Really Interesting New Gene) E3 ligases. These classes differ by the specific mechanisms they use to facilitate the ubiquitin transfer process (Fig.6) [22, 23].

In the HECT family of E3 ligases, ubiquitin is transferred from the E2 enzyme to a cysteine residue located on the E3 HECT domain through the formation of a thioester bond prior to attachment to a substrate [22]. In contrast, the RING E3 domain enables the direct transfer of ubiquitin from the E2 enzyme to the substrate [22]. This classification further describes the diverse mechanisms by which E3 ligases regulate ubiquitination, which is essential for many cellular processes.
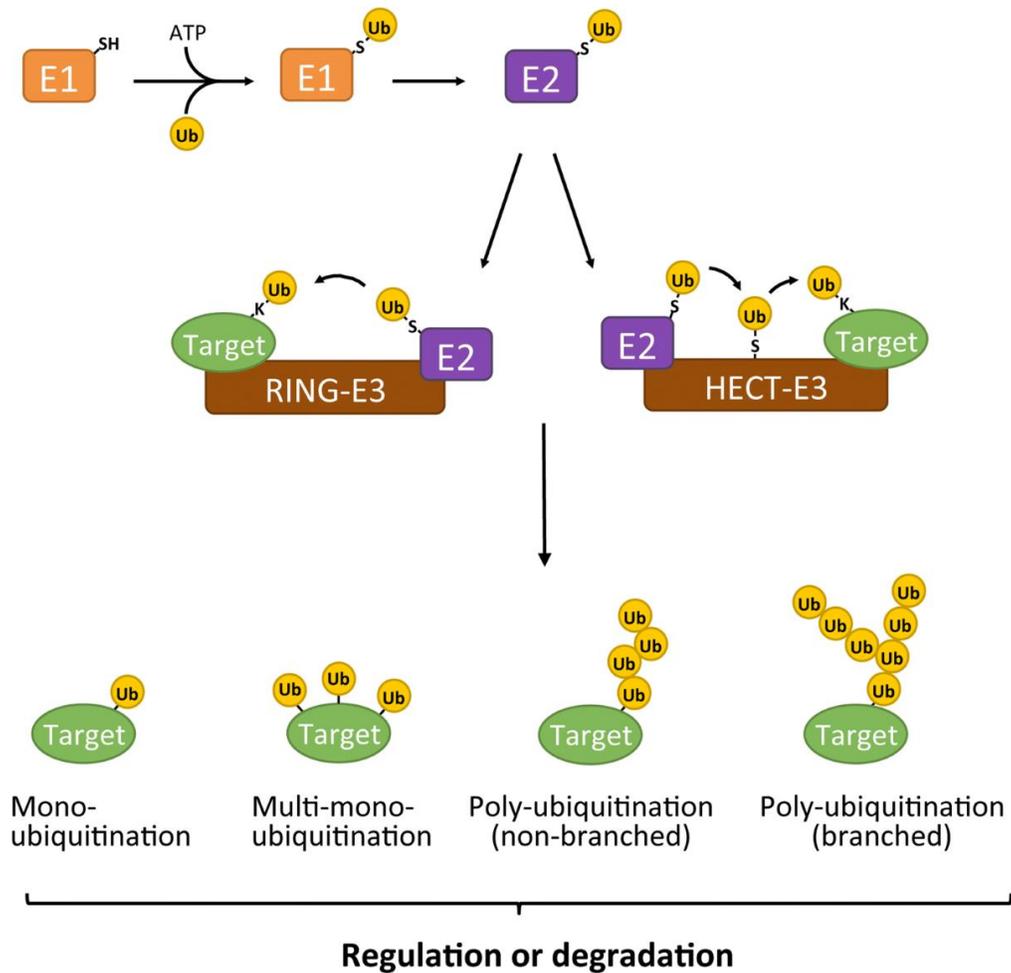
Figure 2-2. The ubiquitination cascade. The initial step involves the ATP-dependent transfer of ubiquitin to an active-site cysteine residue on the E1 ubiquitin-activating enzyme. In the next step, the ubiquitin is transferred from the E1 to the E2 ubiquitin-conjugating enzyme. Once the E2 is charged with ubiquitin, it can associate with the E3 to prepare the ubiquitin transfer to the target protein. A RING-E3 will mediate a direct transfer of ubiquitin from the E2 to the target protein. The substrate ubiquitination by HECT E3s (and RBR ligases, not shown) involves an additional step where the ubiquitin is first chemically bound to an active-site cysteine of the HECT domain before it is attached to the target protein. Target proteins can be modified by mono-, multi-mono-, or polyubiquitin. As ubiquitin has seven internal lysine residues and an N-terminal amino group, all of which can be ubiquitinated, a wide variety of polyubiquitin chains can be formed (not shown). Depending on the type of (poly)ubiquitin modification, either the function/localization of the target protein is changed ("regulation") or the target protein is sent for degradation by the 26S proteasome ("degradation"). (Copied from Sluimer et al, 2018).

Having established the foundational understanding of the TRIM family

proteins, their classification, and their critical role as E3 ubiquitin ligases, it is

important to examine one specific member of the family, TRIM37. The remainder

of the thesis will provide a comprehensive structural and functional overview of

TRIM37, examining its contributions to cellular homeostasis, immune response,

and its emerging role in disease mechanisms, particularly in developmental

disorders and oncogenesis.

2.4 Structural and Functional Overview of TRIM37

The human TRIM37 (also known as MUL) gene is located on chromosome

17q23 and has a full genomic length of approximately 125 kb [24]. The TRIM37

promoter region is G+C-rich and TATA box-less, indicating a unique regulatory

landscape (Fig. 2-3A) [25]. The primary transcript, TRIM37a, is 4.33 kb long and

encodes a protein consisting of 964 amino acids [25]. In addition to TRIM37a, RT-

PCR amplification and in silico analysis of the TRIM37 coding sequence have

revealed several alternative splicing products, with TRIM37b emerging as the most

prominent variant [25]. Although TRIM37b shares the same coding sequence as

TRIM37a, it features a distinct 3' untranslated region (UTR) that likely influences

its stability and translational efficiency [25]. Notably, TRIM37b expression is

substantially lower than that of TRIM37a, except in the testis, where it is

comparably expressed, indicating tissue-specific regulatory mechanisms at play

[25].

TRIM37, like other members of the TRIM family, contains a RING finger domain, a B-box domain 2, and a coiled-coil domain (Fig. 2-3B). Within the TRIM family of proteins, the MATH and DES domains are specific to TRIM37. Located on the C-terminal side of the coiled-coil region, the MATH domain spans 133 amino acids (positions 263–406) and is composed of 6–8 antiparallel β-sheets, enabling it to form both homo- and hetero-oligomeric structures through self-interaction [25]. This domain is shared with other ubiquitin ligases known as TNF receptor-associated factors (TRAFs) and facilitates protein-protein interactions crucial for various signaling pathways. Consequently, the MATH domain, often referred to as the TRAF-C domain, is integral in processes such as proteolysis and ubiquitination, suggesting a shared evolutionary link among proteins within the TRAF and MATH families [11, 26]. In addition to the MATH domain, TRIM37 contains two notable DES-rich sequences, enriched in aspartate, glutamate, and serine residues, located between amino acids 423–555 and 863–943 (Fig.2-3B) [25]. These sequences significantly enhance the solubility of TRIM37, even in the presence of hydrophilic amino acids. Furthermore, TRIM37 also contains a nuclear localization signal (NLS), which is a short sequence motif essential for regulating protein transport into and out of the nucleus. Together, these domains and sequences underscore TRIM37's multifaceted roles in cellular processes and its potential significance in regulating protein dynamics within the cell.

34
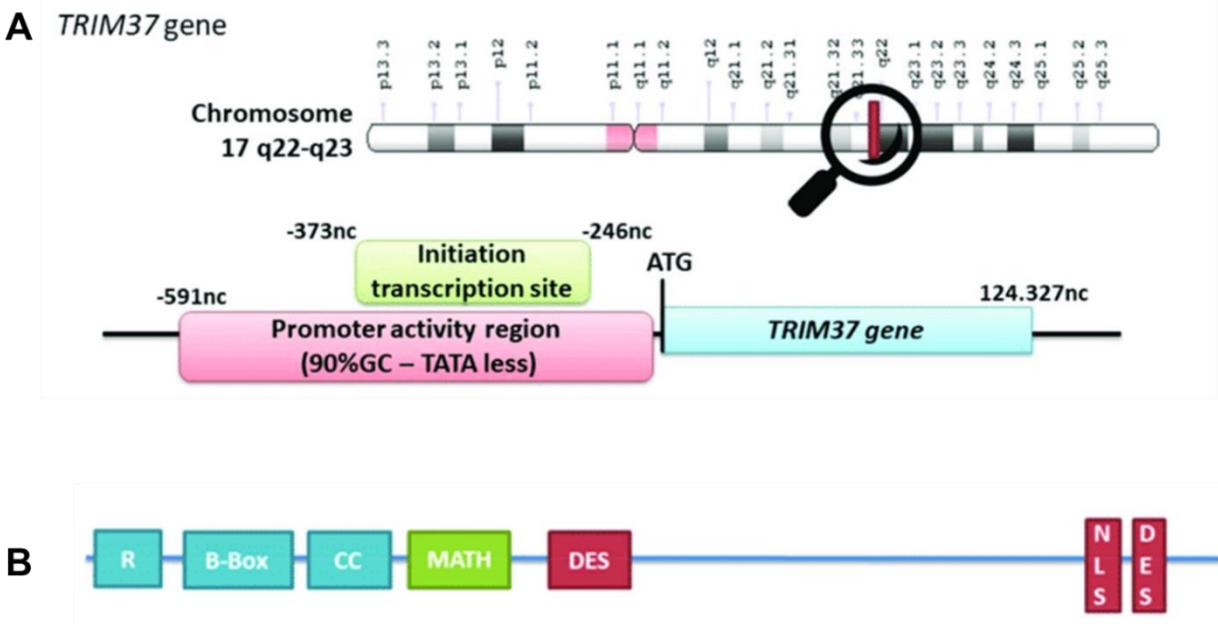
Figure 2-3. The TRIM37 gene. (A) Schematic presentation of the TRIM37 localization and gene structure. (B) TRIM37 protein structure. (Adapted from Brigant et al, 2018).

2.5 Mechanisms of TRIM37 E3 Ubiquitin Ligase Activity

TRIM37, like other members of the TRIM family, has E3 ubiquitin ligase activity, enabling the modification of target proteins through ubiquitination. As an E3 ubiquitin ligase, TRIM37 mono-ubiquitinates histone H2A at lysine 119 (H2AK119), leading to the formation of repressive chromatin and transcriptional silencing [24]. It achieves this by associating with the Polycomb Repressive Complex 2 (PRC2) subunit EZH2, which catalyzes the trimethylation of histone H3 at lysine 26 (H3K26me3), a mark that signals gene silencing [24]. The TRIM37-PRC2 complex further recruits Polycomb Repressive Complex 2 (PRC1) subunit RING finger protein 2 (RNF2), which catalyzes H2A ubiquitination to further

35

compact chromatin, stabilizing the repressive chromatin state and reinforcing gene

silencing at specific gene promoters (Fig.2-4) [24, 27]. The interaction between

TRIM37, PRC2, and PRC1 is crucial for maintaining proper transcriptional

regulation during normal cellular differentiation, development, and identity

maintenance. Dysregulation of TRIM37, whether through mutations or altered

interactions with PRC2, can lead to disrupted gene expression, contributing to

developmental diseases, immune system dysfunction, and oncogenesis.
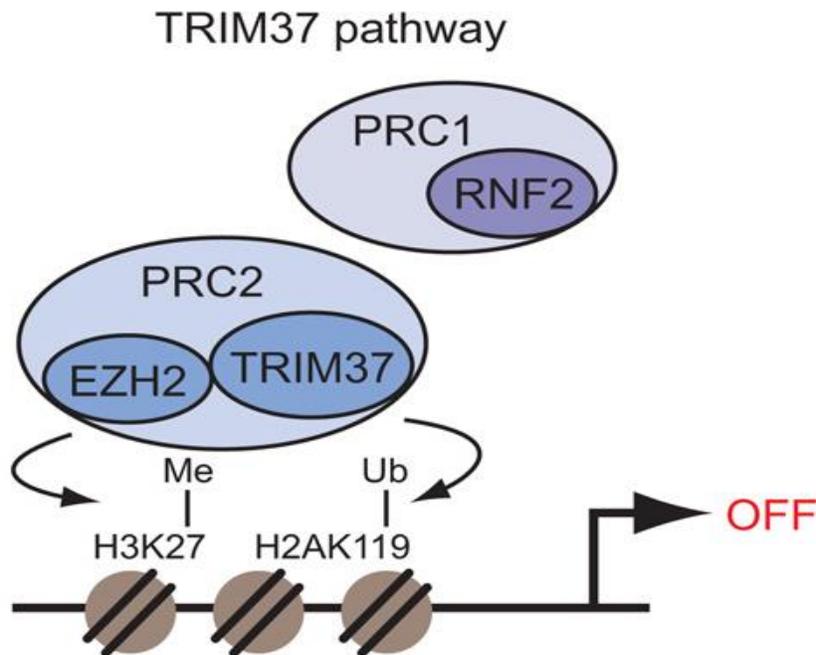
Figure 2-4. Schematic models for TRIM37 pathway for target gene silencing. TRIM37 promotes gene silencing by interacting with the PRC2 subunit EZH2 at specific gene promoters, leading to the deposition of H3K27me3 and H2AK119ub. These modifications induce chromatin compaction and transcriptional silencing. The TRIM37-PRC2 complex further recruits the PRC1 subunit RNF2, which catalyzes H2Aub, further compacting the chromatin and driving gene silencing (Adapted from Bhatnagar et al., 2015).

2.6 Role of TRIM37 in Developmental Diseases

TRIM37 is increasingly recognized for its multifaceted role in developmental disorders, particularly in the context of Mulibrey nanism, a rare genetic condition marked by distinctive features such as short stature, muscle weakness, and various skeletal abnormalities [28, 29]. This disorder is caused by mutations in the TRIM37 gene, disrupting its normal function as an E3 ubiquitin ligase. The impairment of TRIM37's ability to facilitate ubiquitination and degradation of target proteins significantly affects cellular signaling pathways essential for growth and development. Notably, TRIM37 regulates insulin-like growth factor (IGF) signaling, a critical pathway for promoting cell proliferation and differentiation [30]. When TRIM37 function is compromised, the response to growth signals becomes altered, contributing to the growth deficiencies observed in individuals with Mulibrey nanism [31]. Additionally, TRIM37's regulatory role extends to apoptosis, where it helps maintain the balance between cell survival and programmed cell death. Disruption in this balance can reduce functional cells during development, further exacerbating growth and skeletal issues [31].

2.7 TRIM37 in Immune Response Regulation

TRIM37 is increasingly recognized for its significant role in modulating the immune response through its E3 ubiquitin ligase function. One of the key mechanisms by which TRIM37 controls the immune response is by modulating the

37

activity of TRAF6 (tumor necrosis factor receptor-associated factor 6), a protein

that is crucial to the activation of the NF-κB signaling pathway [32]. This pathway,

when activated, leads to the production of pro-inflammatory cytokines such as

TNF-α and IL-6, which are essential for activating immune responses to infections

and injury [32]. By promoting the ubiquitination and degradation of TRAF6,

TRIM37 helps to prevent excessive or prolonged activation of NF-κB, thereby

maintaining a balance in immune signaling. Therefore, abnormal overexpression of

TRIM37 could lead to excessive degradation of TRAF6, reducing the body's ability

to respond to infections and making the body more susceptible to pathogens, as the

NF-κB pathway is crucial for initiating immune defense mechanisms. Furthermore,

TRIM37 has been shown to interact with various pattern recognition receptors

(PRRs), such as Toll-like receptors (TLRs), facilitating the recognition of pathogens

and the subsequent activation of immune cells [33]. This interaction amplifies the

signaling cascades triggered by PRRs and promotes the production of type I

interferons, which are crucial for antiviral defense. Additionally, TRIM37 regulates

apoptosis in immune cells, affecting the balance between cell survival and death

during immune activation. It influences the degradation of key proteins involved in

apoptosis, such as Caspase-8 and IAPs, thereby determining the fate of activated

immune cells [33, 34]. The intricate interplay between TRIM37 and immune

signaling pathways underscores its importance in modulating immune responses to

pathogens and stress. It is a viable target for therapeutic interventions in autoimmune diseases and infections.

2.8 TRIM37 as a breast Cancer oncogene

Increased expression of TRIM37 has been observed in various cancers, such as breast, gallbladder, pancreatic, and colorectal cancers, with evidence indicating that elevated levels of TRIM37 correlate with aggressive tumor characteristics and poorer clinical outcomes [24, 35-37]. Notably, growing evidence has highlighted TRIM37 as a significant player in breast cancer oncogenesis.

TRIM37 was initially described as a novel breast cancer oncoprotein and has since been recognized as a critical player in breast oncogenesis [24, 38, 39]. In our 2014 *Nature* study, we reported that TRIM37 acts as a histone H2A ubiquitin ligase, mono-ubiquitinating histone H2A at Lys119 (H2Aub), a chromatin modification associated with transcriptional repression [24]. In human breast cancer cell lines with amplified 17q23, we found that TRIM37 is upregulated and associated with poor overall survival, highlighting its oncogenic potential in breast cancer [24]. Genome-wide chromatin immunoprecipitation (ChIP) experiments identified many target genes, including multiple tumor suppressors, whose promoters were bound by TRIM37 and enriched for H2Aub [24]. Functionally, overexpression of TRIM37 results in the down-regulation of tumor suppressors and pro-apoptotic genes, substantially increasing tumor growth in mouse xenografts. In

contrast, inhibiting TRIM37 function has been shown to reduce tumor growth in xenograft and patient-derived mouse models [24]. Beyond its role in cancer oncogenesis, TRIM37 is also recognized as a breast cancer predisposition gene that increases the risks of developing the disease [40]. Moreover, ectopic expression of TRIM37 renders non-transformed cells tumorigenic [24]. In addition to its role in driving breast cancer initiation and tumorigenesis, growing evidence suggests that TRIM37 may be integral to the more advanced stages of cancer, particularly in metastasis and chemoresistance. Therefore, the next chapter will explore TRIM37's involvement in these later stages, underscoring its potential as a therapeutic target for both early and late-stage breast cancer.

2.9 References

1.   Sardiello, M., et al., Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. BMC Evol Biol, 2008. 8: p. 225.

2.   Patil, G. and S. Li, Tripartite motif proteins: an emerging antiviral protein family. Future Virol, 2019. 14(2): p. 107-122.

3.   Cao, X., et al., The Role of Tripartite Motif Family Proteins in Chronic Liver Diseases: Molecular Mechanisms and Therapeutic Potential. Biomolecules, 2024. 14(8).

4.   Williams, F.P., et al., Emerging RNA-binding roles in the TRIM family of ubiquitin ligases. Biol Chem, 2019. 400(11): p. 1443-1464.

5.   Jain, K., et al., Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. Epigenetics Chromatin, 2020. 13(1): p. 3.

6.   Musselman, C.A. and T.G. Kutateladze, PHD fingers: epigenetic effectors and potential drug targets. Mol Interv, 2009. 9(6): p. 314-23.

7.   Josling, G.A., et al., The role of bromodomain proteins in regulating gene expression. Genes (Basel), 2012. 3(2): p. 320-43.

8.   Tocchini, C. and R. Ciosk, TRIM-NHL proteins in development and disease. Semin Cell Dev Biol, 2015. 47-48: p. 52-9.

9.   Loedige, I., et al., The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. Genes Dev, 2014. 28(7): p. 749-64.

10.   Nakamura, N., The Role of the Transmembrane RING Finger Proteins in Cellular and Organelle Function. Membranes (Basel), 2011. 1(4): p. 354-93.

11.   Dai, Y., et al., Evolution and Expression of the Meprin and TRAF Homology Domain-Containing Gene Family in Solanaceae. Int J Mol Sci, 2023. 24(10).

12.   Su, Z., K. Dhusia, and Y. Wu, Understand the Functions of Scaffold Proteins in Cell Signaling by a Mesoscopic Simulation Method. Biophys J, 2020. 119(10): p. 2116-2126.

13.   Liu, T., et al., NF-kappaB signaling in inflammation. Signal Transduct Target Ther, 2017. 2: p. 17023-.

14.   Ikeda, K. and S. Inoue, TRIM proteins as RING finger E3 ubiquitin ligases. Adv Exp Med Biol, 2012. 770: p. 27-37.

15.   Cai, C., et al., The RING finger protein family in health and disease. Signal Transduct Target Ther, 2022. 7(1): p. 300.

16.   Lipkowitz, S. and A.M. Weissman, RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. Nat Rev Cancer, 2011. 11(9): p. 629-43.

17.   Dong, Y., et al., The role of E3 ubiquitin ligases in bone homeostasis and related diseases. Acta Pharm Sin B, 2023. 13(10): p. 3963-3987.

18.   Wang, J., et al., The Ubiquitin-Proteasome System in Tumor Metabolism. Cancers (Basel), 2023. 15(8).

19.   Scaglione, K.M., et al., The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. J Biol Chem, 2013. 288(26): p. 18784-8.

20.   Chen, Z.J. and L.J. Sun, Nonproteolytic functions of ubiquitin in cell signaling. Mol Cell, 2009. 33(3): p. 275-86.

21.   Guo, H.J., N. Rahimi, and P. Tadi, Biochemistry, Ubiquitination, in StatPearls. 2025: Treasure Island (FL).

22.   Sluimer, J. and B. Distel, Regulating the human HECT E3 ligases. Cell Mol Life Sci, 2018. 75(17): p. 3121-3141.

23.   Michaelides, I.N. and G.W. Collie, Correction to "E3 Ligases Meet Their Match: Fragment-Based Approaches to Discover New E3 Ligands and to Unravel E3 Biology". J Med Chem, 2023. 66(22): p. 15564.

24.   Bhatnagar, S., et al., TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. Nature, 2014. 516(7529): p. 116-20.

25.   Brigant, B., et al., TRIMming down to TRIM37: Relevance to Inflammation, Cardiovascular Disorders, and Cancer in MULIBREY Nanism. Int J Mol Sci, 2018. 20(1).

26.   Zapata, J.M., V. Martinez-Garcia, and S. Lefebvre, Phylogeny of the TRAF/MATH domain. Adv Exp Med Biol, 2007. 597: p. 1-24.

27.   Buchwald, G., et al., Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J, 2006. 25(11): p. 2465-74.

28.   Kallijarvi, J., et al., TRIM37 defective in mulibrey nanism is a novel RING finger ubiquitin E3 ligase. Exp Cell Res, 2005. 308(1): p. 146-55.

29.   Gu, W., et al., The TRIM37 variants in Mulibrey nanism patients paralyze follicular helper T cell differentiation. Cell Discov, 2023. 9(1): p. 82.

30.   Chen, J., et al., Role of the tripartite motif-containing (TRIM) family of proteins in insulin resistance and related disorders. Diabetes Obes Metab, 2024. 26(1): p. 3-15.

31.   Wang, W., et al., TRIM37 deficiency induces autophagy through deregulating the MTORC1-TFEB axis. Autophagy, 2018. 14(9): p. 1574-1585.

32.   Zhao, L., et al., TRIM37 negatively regulates inflammatory responses induced by virus infection via controlling TRAF6 ubiquitination. Biochem Biophys Res Commun, 2021. 556: p. 87-92.

33.   Ahsan, N., et al., Multipronged regulation of autophagy and apoptosis: emerging role of TRIM proteins. Cell Mol Biol Lett, 2024. 29(1): p. 13.

34.   Li, Z.Q., X. Chen, and Y. Wang, Small molecules targeting ubiquitination to control inflammatory diseases. Drug Discov Today, 2021. 26(10): p. 2414-2422.

35.   Xu, M., et al., TRIM37 promotes gallbladder cancer proliferation by activating the Wnt/beta-catenin pathway via ubiquitination of Axin1. Transl Oncol, 2023. 35: p. 101732.

36.   Do, T.T., et al., TRIM37 Promotes Pancreatic Cancer Progression through Modulation of Cell Growth, Migration, Invasion, and Tumor Immune Microenvironment. Int J Mol Sci, 2022. 23(3).

37.   Zhao, P., et al., Knockdown of Tripartite Motif-Containing Protein 37 (TRIM37) Inhibits the Proliferation and Tumorigenesis in Colorectal Cancer Cells. Oncol Res, 2017. 25(1): p. 115-122.

38.   Bhatnagar, S. and M.R. Green, TRIMming down tumor suppressors in breast cancer. Cell Cycle, 2015. 14(9): p. 1345-6.

39.   Tushir-Singh, J. and S. Bhatnagar, In Vitro Assay to Study Histone Ubiquitination During Transcriptional Regulation. Methods Mol Biol, 2017. 1507: p. 235-244.

40.   Rahman, N., Realizing the promise of cancer predisposition genes. Nature, 2014. 505(7483): p. 302-8.

Chapter 3: Oncogenic TRIM37 Links chemoresistance and metastatic Fate in triple-negative Breast cancer

3.1 Abstract

The majority of clinical deaths in TNBC patients are due to chemoresistance and aggressive metastases, with high prevalence in younger women of African ethnicity. While tumorigenic drivers are numerous and varied, the drivers of metastatic transition remain largely unknown. Here, we uncovered a molecular dependence of TNBC tumors on the TRIM37 network, which enables tumor cells to resist chemotherapeutic and metastatic stress. TRIM37-directed histone H2A mono-ubiquitination enforce changes in DNA repair that rendered TP53-mutant TNBC cells resistant to chemotherapy. Chemotherapeutic drugs triggered a positive feedback loop via ATM/E2F1/STAT signaling, amplifying the TRIM37 network in chemoresistant cancer cells. High expression of TRIM37 induced transcriptomic changes characteristic of a metastatic phenotype, and inhibition of TRIM37 substantially reduced the *in vivo* propensity of TNBC cells. Selective delivery of TRIM37-specific antisense oligonucleotides using anti-folate receptor 1-conjugated nanoparticles in combination with chemotherapy suppressed lung metastasis in spontaneous metastatic murine models. Collectively, these findings establish TRIM37 as a clinically relevant target with opportunities for therapeutic intervention.

3.2 Introduction

Triple negative breast cancer (TNBC) is an aggressive breast cancer subtype that accounts for ~20% of all breast cancer cases with the annual incidence rate in the United States estimated to be 40,000 [1]. TNBC patients are disproportionately associated with the highest frequency of chemoresistance, relapse, and metastasis. Consequentially, the 5-year survival rate of TNBC is 66% relative to 93% for other breast cancer subtypes [1]. Despite the high mortality rate in women worldwide, chemotherapy remains the standard of care for TNBC patients. Although chemotherapy is effective initially in TNBC patients, it is often accompanied by resistance, relapse, and severe side effects. Therefore, new and effective targeted therapies to prevent and ultimately cure TNBC are a clinical priority. While lifestyle, epidemiologic, and cultural factors shape TNBC clinical outcome, the disease etiology is also dependent on biogeographical ancestry [2]. Thus, lack of targeted therapies for TNBC is fraught with multiple challenges attributed to limited understanding of genetic complexities, metastatic biology, and drivers of metastatic traits.

An unresolved question in cancer biology is what drives a primary tumor to become metastatic? This is a clinically relevant question because metastatic, not primary tumors, are fatal. In general, numerous oncogenes and tumor suppressors are genetically or epigenetically altered in cancer and accumulate during

tumorigenesis. But whether drivers of tumorigenesis are also the causal factor of the metastatic transition remains to be addressed. To this end, extensive transcriptomic and genetic scans of evolving carcinomas revealed mutations that were represented in premalignant biopsies but not in tumor biopsies, suggesting divergence of genetic alterations during the transition from primary to regional metastases [3-4]. Additionally, dynamic epigenetic mechanisms are also intimately linked to metastatic transitions [5-6]. For example, TNBC tumors harbor a high frequency of hypermethylated promoters in commonly targetable drivers, such as TP53, BRAF, KRAS, and EGFR [7]. Alterations in epigenetic factors causing neomorphic mutations (e.g., EZH2, DNMT3A) or translocations (e.g., NSD2, MMSET) are also frequent in cancer patients [6]. As such, several small molecules targeting epigenetic regulators have entered clinical trials, for example, Estinosat, Belinostat, and Panobinostat [8]. Understandably, these drug treatments are not mutation-specific and thus, pose a significant toxicity risk to untransformed cells, underscoring the critical need for targeted therapies.

We have originally described tripartite motif-containing protein 36 (TRIM37) as a breast cancer oncoprotein that can epigenetically silence tumor suppressors [9, 10]. Clinically, high-TRIM37 associate with poor overall survival [9]. Mechanistically, TRIM37 mono-ubiquitinates histone H2A at Lys119 (H2Aub) to down-regulate target genes [9]. Functionally, TRIM37 over-expression renders non-

transformed breast cells tumorigenic, and inhibition of TRIM37 function reduces tumor growth [9]. While TRIM37 promotes tumorigenesis, its function in breast cancer metastasis and the therapeutic implications of TRIM37 targeting remain to be demonstrated.

Metastasis is a multistep process that includes pathways regulating the epithelial-mesenchymal transition, infiltration of distant sites, and metastatic growth [11]. Given the majority of TNBC patients receive chemotherapy, the ability to resist therapy-induced cell death is perhaps the first step towards a metastatic phenotype. Indeed, recent evidence revealed synchronized expression of genes involved in surviving the stress of chemotherapy as well as overcoming the natural barriers of metastatic growth. For example, a CXCL1/2 paracrine pathway [12] and MTDH [13] were recently identified with dual functionality in metastasis and chemoresistance. Here, we uncover that TRIM37 alters DNA damage response to prevent therapy-induced cell death, and enforces a transcriptional program favoring metastasis. In particular, we report that selective TRIM37 inhibition in TNBC tumors suppress lung metastases *in vivo*. Together, these data reveal that TRIM37 is a new epigenetic driver of aggressive TNBC biology, which can be targeted to simultaneously increase chemotherapy efficacy and reduce metastasis risk in TNBC patients.

3.3 Materials and Methods

Cell lines and cell culture

MDA-MB-231-luc-D3H2LN-BMD2b (provided by Takahiro Ochiya), MDA MB 231, MDA MB 468 and HCC1806 (provided by Michael J. Lee) cells were maintained in RPMI medium supplemented with 10% fetal bovine serum (FBS, Invitrogen) at 36°C and 5% CO_2. MCF10A, MCF6, MCF10AT, HCC1806RR (provided by Sophia Ran) and TP53−/− MCF10A (provided by David Weber and Michele Vitolo) were cultured as described previously [9, 14, 15]. Cells cultured at the same time were pooled together and then seeded after counting in a 6-well or 10-cm dish. Cells were then subjected, in a random order, to treatment with a control or test different biologics, which included shRNA, sgRNA, vectors, and small molecule inhibitors. Cells were routinely tested for mycoplasma using PlasmoTest kit from (Invitrogen).

Animal Care

NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ and Balb/cJ (Jackson Laboratory) were housed in a specific-pathogen-free facility accredited by the American Association of Laboratory Animal Care. All animal studies were approved (#4112 and #4222) by the Institutional Animal Care and Use Committee.

CRISPR/Cas9 targeting

The Gapdh sgRNAs were cloned in LentiCrispr v2 plasmid (Addgene) and packaged into virus as recommended by the manufacturer. Cells were infected with packaged virus as described previously [9].

Recombinant Antibody Cloning

Farletuzumab (anti-human FOLR1) and LK26 (anti-mouse FOLR1) antibodies were cloned, engineered, expressed, and purified as described previously [16]. To generate an antibody conjugate platform for nanoparticles linkage, a novel linker sequence [(X)3Cys(X)3 amino acid sequence] was engineered in continuation of the carboxy terminal of heavy chain (called Fc-Linkered). The knob chain is exactly similar to hole chain except for the presence of this unique cysteine residue.

Recombinant Antibody Expression

Free style CHO-S cells (Invitrogen) were cultured and maintained according to supplier's recommendations (Life technologies). A ratio of 2:1 (light chain, VL: heavy chain, VH) DNA was transfected using 1 mg/ml polyethylenimine and cultured at 36°C. After 24 hrs., transfected cells were cultured at 32°C for additional 9 days. Cells were fed every 2nd day with 1:1 ratio of Tryptone feed and CHO Feed B and antibodies were purified as described previously [16]. An Autodesk Inventor Professional 2020 was used to draw the design of antibody.

Nanoparticle synthesis and packaging

Cholesterol, 1,2-dioleoyl-3-trimethylammonium-propane chloride salt (DOTAP), 1,2-dioleoyl-sn-glycero-3-phosphate (DOPA), and 1,2-distearoryl-sn-glycero-3-phosphoethanolamine-N-[maleimide-(polyethyleneglycol-2000)] ammonium salt (DSPE-PEG2000-Maleimide) were purchased from Avanti Polar Lipids, Inc. (Alabaster).

TRIM37-ASO (IDT) and control-ASO (IDT) loaded bilayer nanoparticles were prepared as previously described with modification [17]. Briefly, 150mM CaCl2 with 0.5mM ASO were dispersed in Cyclohexane/Igepal CO-520 (60:30 v/v) solution to form water-in-oil reverse micro-emulsion (Solution A). The phosphate phase was prepared by mixing 1.5 µM NaHPO4 (pH=9.0) and 6.9 µM DOPA in Cyclohexane/Igepal CO-520 (60:30 v:v) solution (Solution B). Calcium phosphate-alginate (CaP) core was prepared by mixing Solutions A and B. For assembly of outer leaflet, CaP core was mixed with 0.5 µM DOTAP/Cholesterol (1:1), and 0.15 µM DSPE-PEG-2000-Maleimide. CaP-bilipids nanoparticles were mixed with 0.56 µM engineered Farletuzumab solutions and incubated overnight at 4°C. Finally, CaP-bilipids nanoparticles were sterile filtered for subsequent experiments. An Autodesk Inventor Professional 2020 was used to draw the design of smart nanoparticle.

NSG tumor xenograft studies

For indicated cell lines, weight and aged matched female NSG mice were injected subcutaneously with 2x106 TNBC cells in their right flank as described previously [9]. For *in vivo* Dox-induced TRIM37 upregulation studies, mice bearing ~200 mm3 HCC1806 tumors weight matched animals were randomly assigned into groups and injected with 2 mg/kg Dox. Tumors were harvested 24 hrs. following Dox treatment. For smart nanoparticles efficiency *in vivo*, mice bearing ~200 mm3 231-2b tumors weight matched animals were randomly assigned into groups and injected with control or smart nanoparticles. At the endpoint, RNA and protein lysates were prepared from isolated tumors.

Bioluminescent imaging (BLI)

150 mg/kg luciferin (Perkin Elmer Inc.) was administered to mice intraperitoneally and mice were imaged as described previously [16].

Spontaneous metastatic tumor studies

2×10^6 4T1 or 231-2b cells were injected into the inguinal mammary fat pad or right flank of 6 to 8-week-old female Balb/cJ or NSG mice and tumor growth was monitored as previously described [9]. Where indicated, primary tumors were resected and animals were allowed to develop lung metastases. Mice bearing TNBC tumors were weight matched and randomly assigned into groups that received 1.2 mg/kg smart or control nanoparticles at the indicated times. The lung metastases in the animals were monitored by BLI. At the termination of the experiment, animals

were euthanized and indicated tissues were harvested and processed for histological examination and immunohistochemical staining or for qRT-PCR analysis. Metastatic burden was calculated either as the number of visible metastatic lesion in each organ or as the relative luminescence signal from gross organ tissue.

Experimental metastasis *in vivo*

$3x10^5$ 231-2b cells were inoculated directly into the left cardiac ventricle. Metastatic growth was monitored using BLI. Lungs, liver, femurs and brains were harvested post-mortem and processed for histological examination and gross analysis.

Bioinformatic analysis

cBioPortal was used to obtain TCGA expression z-scores for genes in Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort [18]. FireBrowse (Broad Institute) was used to obtain normal breast tissue expression z-scores. Cooccurrence and mutual exclusivity analysis were performed with cBioPortal. Hazard ratios were assessed with Cox proportional hazard model for which TNBC and non-TNBC patients were stratified into high- and low-TRIM37 as third and first quartile. For Kaplan-Meier analysis, the p-value was calculated with log-rank test. For boxplot analysis, TRIM37-regulated genes were stratified according to TRIM37 expression using z score thresholds ($z>0.5$, $z<-0.5$).

Statistical analysis

All experiments were performed at least in triplicate and the results presented are the mean of at least three different biological replicates. The comparisons between the two groups were done by unpaired t-test; comparisons between multiple treatment groups were done by one-way or two-way ANOVA with indicated multiple comparisons post hoc tests. The enrichment of genes positively correlating with TRIM37 were calculated with two-tailed Fisher's exact test. The distributions in correlation between GO terms was calculated by Kolmogorov–Smirnov test. All statistical analyses were performed using R/Bioconductor (version 2.15.2).

Table 3-1: List of primers and shRNA used in this study

qRT-PCR	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
TRIM37	AACAGAGCGTGGAGAGCATT	CTTCTGCCCAACGACAATTT
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
NBS1	ATGCACTCACCTTGTCATGG	ACGTCCACAAATGAGTGCAC
RAD51C	CGAGCTTAGCAAAGAAGTTGGG	TGCTCAAGAAGTTCCAGTGC
KU70	TTTCAAGCCGTTGGTACTGC	ACAATGCTGCAACCTCCTTC
KU80	TGCACTAATCGTGAGCATGG	TTTTGCTGAATCGGCTGCTG
HK2	TGGACAAAAGGCTTCAAGGC	TTCACCACAGCAACCACATC
GPI	TTCCAGAATTGCAGCAATG	AAGCTGAAGTGGTTGAAGCG
TGM2	AGATGTCAACCCCAAGTTCCTG	ATCGTTGCAGTTGACCATGC
VEGFA	GGCAAAAACGAAAGCGCAAG	CAAATGCTTCTCCGCTCTGAG
PDGFRB	TGAACGTGGTCAACCTGTTG	AGGAAGGTGTGTTTGTTGCG
IGFBP3	AAAAAGCAGTGTCGCCCTTC	ACTTGCTCTGCATGCTGTAG
EGLN3	TGCTGAAGAAAGGGCAGAAG	ACAAGGCCAGCAGATTTCAG
PPFIA4	ATTGGTGGAGTTGGGCAATG	TTACGACATGGCCAACAGTG
REDD1	AAGACACGGCTTACCTGGATG	AAGTGTTCATCCTCAGGGTCAC
SLCO2A1	ACATGATGGTTCTGCGTGTG	AGCGCATCAACAAGAACTGC
S100A4	AGAACTAAAGGAGCTGCTGACC	TGTTGCTGTCCAAGTTGCTC
STC1	ACAGCAAGCTGAATGTGTGC	ATCACATTCCAGCAGGCTTC
VLDLR	TGACATTGCTGCCCAGAAC	ACCCAATCAACAGCAATGGC
SHH	ACTCACCCCCAATTACAACCC	TGGCCAAAGCGTTCAACTTG
L1CAM	TGGCTGCCAATGACCAAAAC	AAGGTCACCCTGGAACCTTTC

BMI1	CGGGAGCTGCTTGGGTCAAGTG	CAGCCTAACACCCACCTTGGCG
EZH2	CAGGCACAGTGGCTCACGTCTG	GTTGGCCAGGCTGGTCTCGAAC
KISS1	TGGTTTCTTGGCAGCTACTG	ATTCTAGCTGCTGGCCTGTG
BRMS1	ATGGTGGGATGACAAACTGC	ATGTATGGGCCAGAAACCAGAG
ChIP		
GAPDH	CCCATGTTCGTCATGGGTGT	GGAGCGTGTCCATAGGGTG
FRA3B	TGTTGGAATGTTAACTCTATCCCAT	ATATCTCATCAAGACCGCTGCA
TRIM37	ACTTAACTACGGGTGTGGCTC	CTGGCGTACTGGTGGAAAGC
ACTIN	AGGGACTCTAGTGCCCGACAC	CCCACCTCCACCCTACCTGC
BRMS1	TAACTAGCTGTGTGGCCTTGAG	ATCTGAGCCTCCCAACAGTTC
KISS1	TCCCATTTCGTCCTCCTTGC	TCCTTCCTGCTTCCCTTCTTTC
Cloning		
TRIM37	TAAAGCGGCCGCCATGGATGAACAGAGCGTGGAGAG	CCGCGGATCCTTATCTTCCACTATTTTCATCTGTATTG
shRNA (Dharmacon)		**Catalogue #**
TRIM37 shRNA#1	TTCACTGGTAAAGTCTGGTGG	TRCN0000034209
TRIM37 shRNA#2	ATGACCTCCTTTAGCATTAGC	TRCN0000034211
BMI1 shRNA#1	TTCTCCAGGTATAAATGTAGG	TRCN0000020154
BMI1 shRNA#2	TTCCGATCCAATCTGTTCTGG	TRCN0000020158
BRMS1 shRNA#1	TTTAGCTCCGAGAACTGCTTC	TRCN0000038894

BRMS1 shRNA#2	AAGCATGTACACGATGTATGG	TRCN0000038896
KISS1 shRNA#1	AAGGAGTTCCAGTTGTAGTTC	TRCN0000059063
KISS1 shRNA#2	TTCCAGTTGTAGTTCGGCAGG	TRCN0000059065
EZH2 shRNA#1	GCAGCTGGTGAGAAGGCAATA	TRCN0000040073
EZH2 shRNA#2	TGAAGCTAAGGCAGCTGTTT	TRCN0000040075
E2F1 shRNA#1	AAAGGGCCGAAAGTGCAGTTA	TRCN0000000249
E2F1 shRNA#2	CAAAGTCACAGTCGAAGAGGT	TRCN0000000253
STAT1 shRNA#1	ACTGTGTTCATCATACTGTCG	TRCN0000004266 TRCN0000004268
STAT1 shRNA#2	TTCTCGTCCTGATACTTTGGG	TRCN0000004268 TRCN0000020839
STAT3 shRNA#1	ATCTCCTGACCTTATGATCCG	TRCN0000020839
STAT3 shRNA#2	AAGTGGCATGTGATTCTTTGC	TRCN0000020843

3.4 Results

TRIM37 associates with double strand break (DSB) repair machinery in TNBC

Our analysis of the METABRIC cohorts revealed that TRIM37 associates with DNA repair genes (Fig.3-1A). Double strand break (DSB) repair is one of the major pathways to repair damaged DNA in cancer cells and its kinetics predicts resistance to therapy [21-22]. We therefore limited the analysis to repair proteins that participate in homologous recombination (HR) and non-homologous end-joining repair pathways (NHEJ), the two subtypes of DSB repair. RAD51C, XRCC5 (Ku80), RNF8, XRCC6BP1 (Ku60 binding protein), RNF168 and MRE11 were among the DSB genes whose expression most strongly associated with TRIM37 (Fig.3-1A). RAD51-associated proteins generally promote HR repair [23]. XRCC5 and XRCC6BP1 are involved in NHEJ repair pathway [23]. MRE11 forms complex with NBS1 and RAD50 to regulate response at DSB [24]. RNF8 and RNF168 are E3 ligases that promote binding of DSB genes [23]. Additionally, TRIM37 also correlated with the overexpression of a family of other DSB factors, including RAD51AP1, SFR1, DDX1, RAD51, ERCC1 and CHEK2. Together, these findings identified 28 DSB genes that significantly correlated with TRIM37 in breast cancer patients. Furthermore, the analysis of METABRIC cohorts stratified by breast cancer subtypes revealed statistically significant correlation between

TRIM37 and DSB genes in TNBC patients but not in non-TNBC patients or histologically normal breast tissue adjacent to tumor (NAT, Fig.3-1B). As shown in Fig.3-1C, the hazard ratio (HR) was ~2.3-fold higher for TNBC patients with high- compared to low-TRIM37 expression, linking TRIM37 to poor prognosis in TNBC patients. In contrast, no significant association between TRIM37 levels and survival was observed in non-TNBC patients (Fig.3-1D).

We next performed a series of functional experiments to determine the molecular mechanisms underlying the biological activity of TRIM37 in DSB repair using several human TNBC cell lines (HCC1806, MDA MB 468 and MDA MB 231). We first asked whether TRIM37 was physically associated with DSB genes. To test this idea, HCC1806 whole cell extract was fractionated by sucrose gradient sedimentation, and individual fractions were analyzed for TRIM37 and a representative subset of DSB proteins identified in Fig.3-1A. Results demonstrated that TRIM37 co-sediments with XRCC5, XRCC6, RAD51c, MRE11, and NBS1. Physical interactions between TRIM37 and DSB proteins could be confirmed by co-immunoprecipitation in MDA MB 468 cells treated with doxorubicin (Dox), a first-line chemotherapeutic agent (Fig.3-1E). No significant changes in the expression of DSB genes were observed in HCC1806 cells expressing TRIM37-specific short hairpin RNA (shRNA), excluding TRIM37-mediated transcriptional regulation of these DSB genes.

Prompted by these findings, we interrogated TRIM37 recruitment to DSB using two independent experimental systems. We first enzymatically-induced DSB at Gapdh using sequence-specific guide RNA (sgRNA) to promote Cas9-nuclease binding. Chromatin immunoprecipitation (ChIP) assay confirmed TRIM37 as well as H2Aub enrichment at the DSB in Gapdh (Fig.3-1F). Next, we analyzed recruitment of TRIM37 and BMI1, a component of the polycomb complex that participates in DSB repair [25], to the endogenous fragile site, FRA3B. Consistently, ChIP analysis confirmed TRIM37 and BMI1 binding to FRA3B following Dox treatment (Fig.3-1G). In marked contrast to the control cells, knockdown of TRIM37 substantially decreased H2Aub enrichment as well as DSB proteins binding to FRA3B in Dox-treated MDA MB 468 cells (Fig.3-1H). As expected, repair efficiencies of NHEJ and HR pathways were significantly reduced in TRIM37-knockdown cells compared to control cells (Fig.3-1I). Together, these results demonstrate that TRIM37 interacts with DSB repair factors and functionally contributes to the repair of therapy-induced DNA damage.
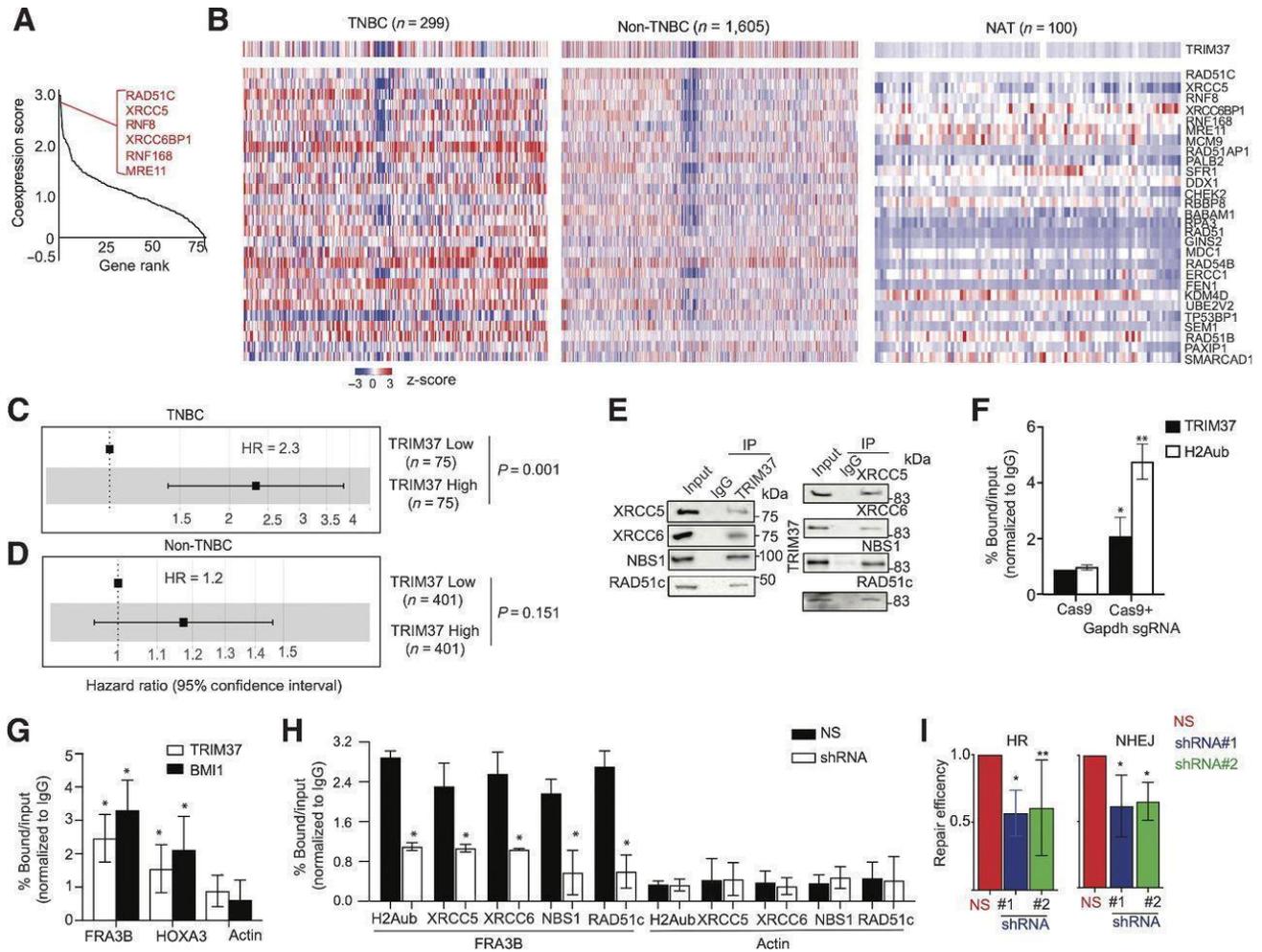
Figure 3-1. TRIM37 associates with DSB repair proteins in TNBC. (A), Gene ranking of DNA repair genes according to cooccurrence with high TRIM37 in patients with breast cancer. The inset shows genes that have the highest correlation with TRIM37. (B), Heat map for the expression of TRIM37 and 28 DSB repair genes in patients with breast cancer stratified by TNBC (n = 299), non-TNBC (n = 1605), and NAT (n = 100). n, number of samples. C and D, Forest plot of HR in TNBC (C) and non-TNBC (D) patients stratified for high- and low-TRIM37 expression using METABRIC cohorts, P values were computed using the log-rank test. (E), Immunoblot monitoring of XRCC5, XRCC6, NBS1, RAD51C, and TRIM37 in protein complexes pulled down by either anti-TRIM37 (left), the indicated DSB proteins (right), or an IgG control. Input, ~1%–5% of whole cell lysates. (F), ChIP monitoring TRIM37 and H2Aub binding at Gapdh in MDA-MB-468 cells expressing either Cas9 alone or with Gapdh site-specific sgRNA, P values were computed using the unpaired t test. (G), ChIP monitoring TRIM37 and BMI1 binding at FRA3B, HOXA3, and Actin in MDA-MB-468 cells treated with Dox, P values were computed using the unpaired t test. (H), ChIP monitoring H2Aub, XRCC5, XRCC6, NBS1, and RAD51c binding at FRA3B and Actin in MDA-MB-468 cells expressing a nonsilencer (NS) or TRIM37 shRNA, P values were computed using the unpaired t test. (I), HR (left) and NHEJ (right)-mediated DSB-repair activity in MDA-MB-468 cells expressing control or TRIM37 shRNA (#1, #2), P values were computed using the unpaired t test. Error bars indicate SD and range of at least three biological replicates. *, p < 0.05; **, p < 0.01; ***, p<0.001; ns, not significant.

TRIM37 catalyzed H2Aub is required for its function in chemoresistance

We next directly examined the impact of TRIM37-knockdown on chemotherapy-induced DNA damage and clonogenic growth. Knockdown of TRIM37 resulted in ~6-fold increase in median tail length in comet assay (Fig.3-2A) and ~5-fold increase in the median nuclear coverage of phosphorylated histone H2AX (γ-H2AX, Fig.3-2B) in Dox-treated cells. TRIM37-knockdown in TNBC cell lines also markedly increased Caspase 3 activity and PARP cleavage (Fig.3-2C), hallmarks of cell death. By contrast, Dox-treated MCF6, a hormone receptor positive breast cancer cell line, did not augment PARP cleavage or Caspase 3 activity (Fig.3-2C), suggesting TRIM37 function in chemoresistance was limited to TNBC cells. As expected, knockdown of TRIM37 sensitized MDA MB 468 cells to chemotherapeutic stress without affecting proliferation as indicated by substantially decreased clonogenic growth relative to the control cells (Fig.3-2D).

Given ubiquitin is critical for DSB factor recruitment to damaged DNA [26], we next asked whether TRIM37-catalyzed H2Aub is required for its function in chemoresistance. To test this idea, we ectopically expressed either wild type TRIM37 (TRIM37) or catalytically-dead TRIM37 (TRIM37(C18R)) in MCF10AT, a premalignant K-RAS transformed triple negative breast cell line. While Dox-treatment induced significant enrichment of H2Aub at FRA3B in TRIM37 overexpressing cells, TRIM37(C18R) failed to promote H2Aub enrichment at

FRA3B (Fig.3-2E). Consequentially, Dox-treatment of TRIM37(C18R)-expressing

cells caused substantially longer comet tails (Fig.3-2F) as well as increased Caspase

3 activity (Fig.3-2G) and PARP cleavage (Fig.3-2H) relative to TRIM37-expressing

cells. Finally, substantially fewer colonies were observed for cells expressing

TRIM37(C18R) compared to TRIM37 following chemotherapeutic stress (Fig.3-

2I).
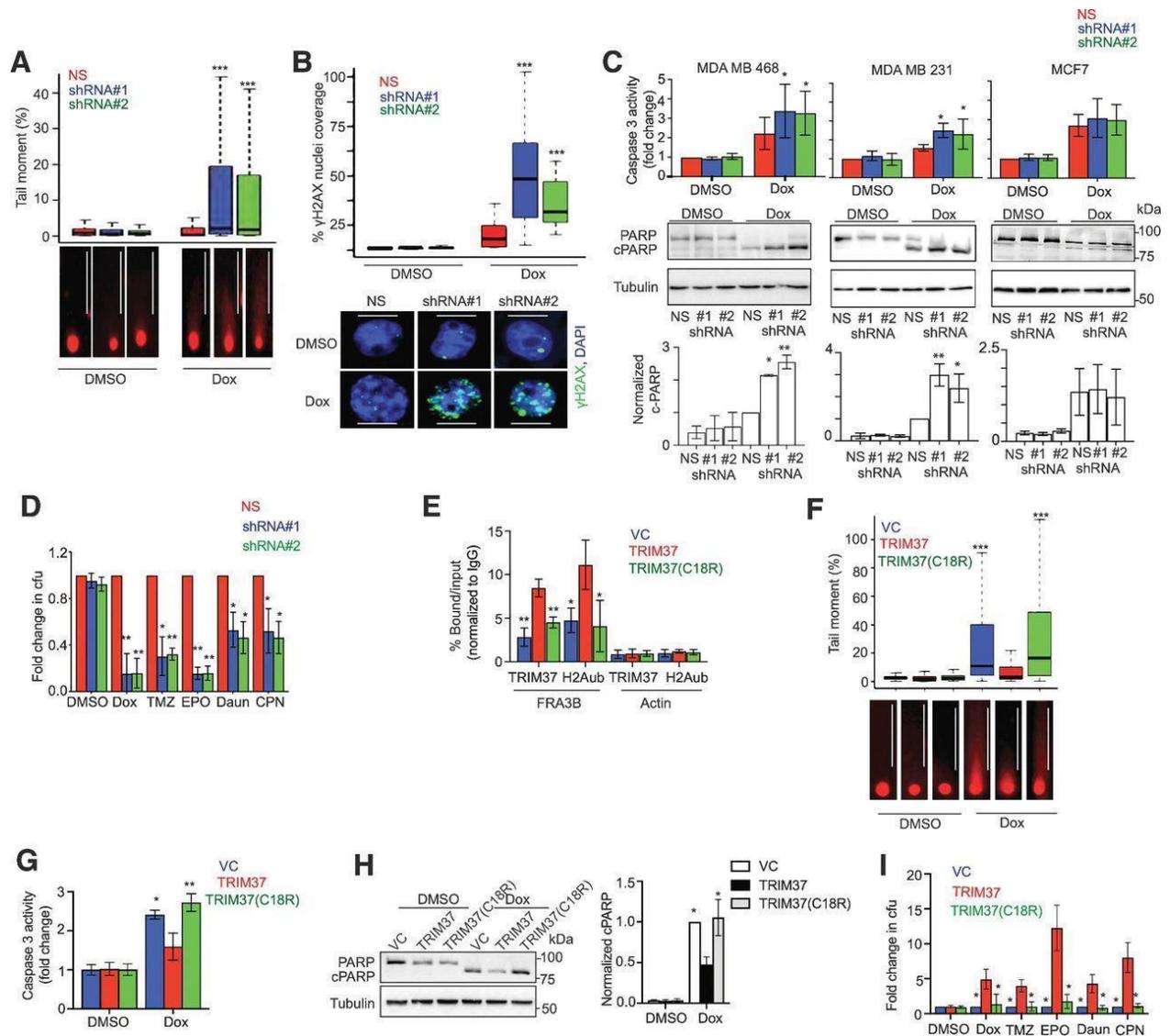
Figure 3-2. TRIM37-catalyzed H2Aub is required for chemoresistance in TNBC. A, Top, tail moment in DMSO or Dox-treated MDA-MB-468 cells expressing nonsilencer (NS) or TRIM37 shRNA (#1, #2). Bottom, representative images of the tails for each group are shown. Scale bars, 100 μm. B, Top, quantification of γH2AX foci in MDA-MB-468 cells expressing nonsilencer or TRIM37 shRNA (#1, #2) following treatment with Dox. Bottom, representative immunofluorescence images of γH2AX foci (green) in Dox-treated MDA-MB-468 cells expressing nonsilencer or TRIM37 shRNA (#1, #2). DAPI (blue) stains the nucleus. Scale bars, 50 μm. C, Caspase-3 activity assay (top) and immunoblot for PARP and cleaved PARP (c-PARP; middle) in DMSO or Dox-treated MDA-MB-468, MDA-MB-231, and MCF6 cells expressing nonsilencer or TRIM37 shRNA (#1, #2). Bottom, quantification of PARP cleavage relative to total PARP is shown. D, Quantification of the fold change in chemotherapeutic drug-resistant colonies obtained for MDA-MB-468 cells expressing either nonsilencer or TRIM37 shRNA (#1, #2) by a clonogenic assay. Cells were treated with DMSO, Dox, temozolomide (TMZ), etoposide (EPO), daunorubicin (Daun), and cisplatin (CPN). Results were normalized to the colony-forming unit (cfu) for DMSO. E, ChIP monitoring TRIM37 and H2Aub binding at FAR3B and Actin in MCF10AT cells expressing vector control (VC), TRIM37, or mutant TRIM37 [TRIM37(C18R)]. F, Top, tail moment in DMSO or Dox-treated VC-, TRIM37-, or TRIM37(C18R)-expressing MCF10AT cells. Bottom, representative images of the tail moment for each group are shown. Scale bars, 100 μm. G and H, Caspase-3 activity assay (G) and immunoblot for PARP and c-PARP (H, left) in DMSO or Dox-treated VC, TRIM37, or TRIM37(C18R)-expressing MCF10AT cells. Right, quantification of PARP cleavage relative to total PARP is shown. I, Quantification of the fold change in chemotherapeutic drug resistant colonies obtained for VC, TRIM37, or TRIM37(C18R) by clonogenic assay. Cells were treated with DMSO, Dox, temozolomide, etoposide, daunorubicin, and cisplatin. Results were normalized to the cfu for DMSO. Error bars indicate SD and range of at least three biological replicates. P values were computed using the unpaired t test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant.

TRIM37 promotes chemoresistance in the absence of functional p53

Wild type TP53 represents a barrier to chemoresistance by altering DSB repair responses, activating checkpoints and the stress responses [27]. Strikingly, TNBC tumors frequently harbor disrupting TP53 mutations, which primarily cause the loss of its wild type function [28]. Surprisingly, analysis of representative TRIM37-associated DSB genes showed a striking correlation between TRIM37 and DSB genes expression in TNBC patients carrying mutant TP53 but not in wild type TP53 (Fig.3-3A). Consistently, TP53 mutant, but not wild type, TNBC patients

with high-TRIM37 were at ~2.3-fold higher risk of death relative to low-TRIM37 (Fig.3-3B–C).

To investigate the relationship between p53 and TRIM37, we transiently expressed wild type p53 in MDA MB 468 (carries transcriptionally inactive p53 R263H) and HCC1806 (carries p53 T256Kfs*90 deletion) cells. For each cell line, p53-reconstituted TNBC cells showed significantly higher PARP cleavage (Fig.3-3D) and Caspase 3 activity (Fig.3-3E) compared to the control cells following Dox-treatment. A clonogenic assay confirmed that p53 over-expression sensitized MDA MB 468 (~4-fold) and HCC1806 (~2-fold) cells to Dox despite high levels of TRIM37 (Fig.3-3F). Reciprocally, ectopic expression of TRIM37 in genetically ablated p53 null (p53−/−) MCF10A cells [14] substantially reduced PARP cleavage, and Caspase 3 activity relative to empty vector (Fig.3-3G–H). As expected, TRIM37-expressing p53−/− MCF10A showed an ~4-fold increase in colony formation in comparison to control cells (Fig.3-3I). In summary, consistent with previous results for MDM2, KRAS, and ARID1 A [29], we find that TRIM37 requires loss of p53 to drive the chemoresistant phenotype in TNBC cells.
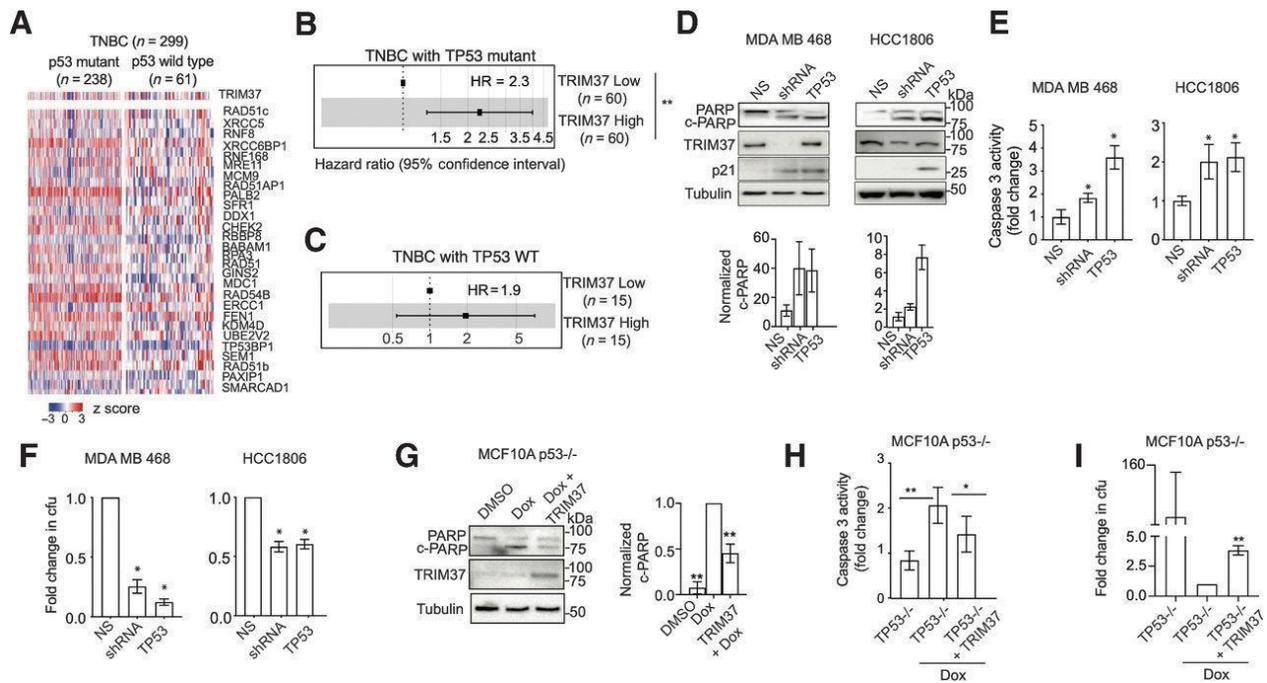
Figure 3-3. TRIM37 reduces cytotoxicity of chemotherapy in the absence of functional p53. A, Heat map for expression of TRIM37 and DSB genes in patients with TNBC stratified by TP53 status [wild-type (n = 238) and mutant (n = 61)]. n, number of patients. B and C, Forest plot of HR for patients with TNBC with mutant TP53 (B), or wild-type TP53 (C) stratified for high- and low-TRIM37 expression. D, Immunoblots in Dox-treated MDA-MB-468 and HCC1806 cells expressing nonsilencer (NS), TRIM37 shRNA, or TP53. Tubulin was the loading control. Bottom, quantification of c-PARP relative to total PARP. E, Caspase-3 activity assay in Dox-treated MDA-MB-468 and HCC1806 expressing nonsilencer, TRIM37 shRNA, or TP53. F, Quantification of the fold change in colony-forming unit (cfu) for Dox-treated MDA-MB-468 and HCC1806 cells expressing nonsilencer, TRIM37 shRNA, or TP53 by clonogenic assay. G, Immunoblots in Dox-treated p53−/− MCF10A cells expressing TRIM37. Tubulin was the loading control. Right, quantification of c-PARP relative to total PARP. H, Caspase-3 activity assay in Dox-treated p53−/− MCF10A cells expressing either vector control or TRIM37. I, Quantification of the fold change in cfu for p53−/− MCF10A cells expressing TRIM37 plated after Dox treatment by clonogenic assay. Error bars indicate SD and range of at least three biological replicates. P values were computed using the unpaired t test. *, $p < 0.05$; **, $p < 0.01$; ***, $p<0.001$; ns, not significant.

Chemotherapy amplifies a TRIM37 survival axis in TNBC

Most TNBC patients receive chemotherapy, which is effective in early stages

of the disease but ~30–50% patients develop resistance [1]. While the exact

mechanisms of chemoresistance remain to be understood, chemotherapeutic drugs

often induce genomic and transcriptomic reprogramming of resistant signatures [30], including alterations in DNA repair capacity [31]. Previous studies have suggested that accumulation of such changes accompany selection and expansion of resistant TNBC cells [32]. We therefore analyzed the expression of TRIM37 in MDA MB 468, HCC1806 and MDA MB 231 cells following chemotherapy. Surprisingly, we found that TRIM37 is upregulated in all the three TNBC cell lines tested in a time-dependent manner (Fig.3-4A). In contrast, no significant increase in TRIM37 was observed in p53 wild type MCF6 or MCF10A, an immortalized breast epithelial cell (Fig.3-4A). The analysis of TRIM37 upregulation kinetics in MCF6 and MCF10a revealed quick and robust p53 activation following Dox treatment, supporting our previous findings that p53 overrides TRIM37 function in chemoresistance (Fig.3-3). Indeed, ectopic expression of p53 in MDA MB 468 cells obliterated Dox-induced TRIM37 upregulation. As expected, TNBC cells treated with additional chemotherapeutic drugs also increased TRIM37 expression post-treatment (Fig.3-4B). Finally, increased TRIM37 level in xenograft tumors following dox-treatment revealed therapy-induced transcriptional upregulation of TRIM37 *in vivo* (Fig.3-4C–D).

We next sought to determine the mechanistic basis for chemotherapy-induced burst in TRIM37 levels in TNBC tumors. A previous study identified TRIM37 association with ataxia-telangiectasia-mutated (ATM) kinase, a DNA damage

sensor [33]. Moreover, TRIM37 promoter harbors regulatory elements for STAT

and E2F1, downstream effectors of ATM kinase. We therefore investigated the

potential role of ATM signaling in the transcriptional regulation of TRIM37 by

utilizing small molecule inhibitors of either ATM or its downstream effectors, E2F1

and JAK (Fig.3-4E). JAK is a tyrosine kinase that phosphorylates STAT. qRT-PCR

analysis revealed that pharmacological inhibition of ATM or its downstream

effectors blocks TRIM37 upregulation following Dox-treatment (Fig.3-4F).

Similarly, E2F1- and STAT1/3-knockdown abolished Dox-induced TRIM37

upregulation, whereas their overexpression significantly increased TRIM37 levels.

ChIP analysis confirmed Dox-induced STAT1, STAT3, and E2F1 recruitment to

TRIM37, which substantially decreased following pharmacological inhibition of

JAK or E2F1 activation (Fig.3-4G–I). Consequently, inhibition of ATM signaling

in MDA MB 468 cells induced significantly higher DNA damage and cell death

relative to control cells as determined by γ-H2AX and PARP cleavage (Fig.3-4J).

Finally, to clinically validate chemotherapy-induced burst in TRIM37, we analyzed

TRIM37 expression in a panel of matched pre- and post-neoadjuvant

chemotherapy-treated tumor biopsies from TNBC patients. qRT-PCR analysis

showed that chemotherapy treatment increased TRIM37 expression in ~82% of

TNBC tumors carrying a mutation in TP53 (n=11, Fig.3-4K). Consistent with

TNBC cellular models, no significant change in TRIM37 was observed in TP53-

wild type TNBC tumors (n=6; Fig.3-4K). Collectively, our results show that

chemotherapeutic stress increases TRIM37 expression in TNBC tumors in an

ATM-dependent manner.
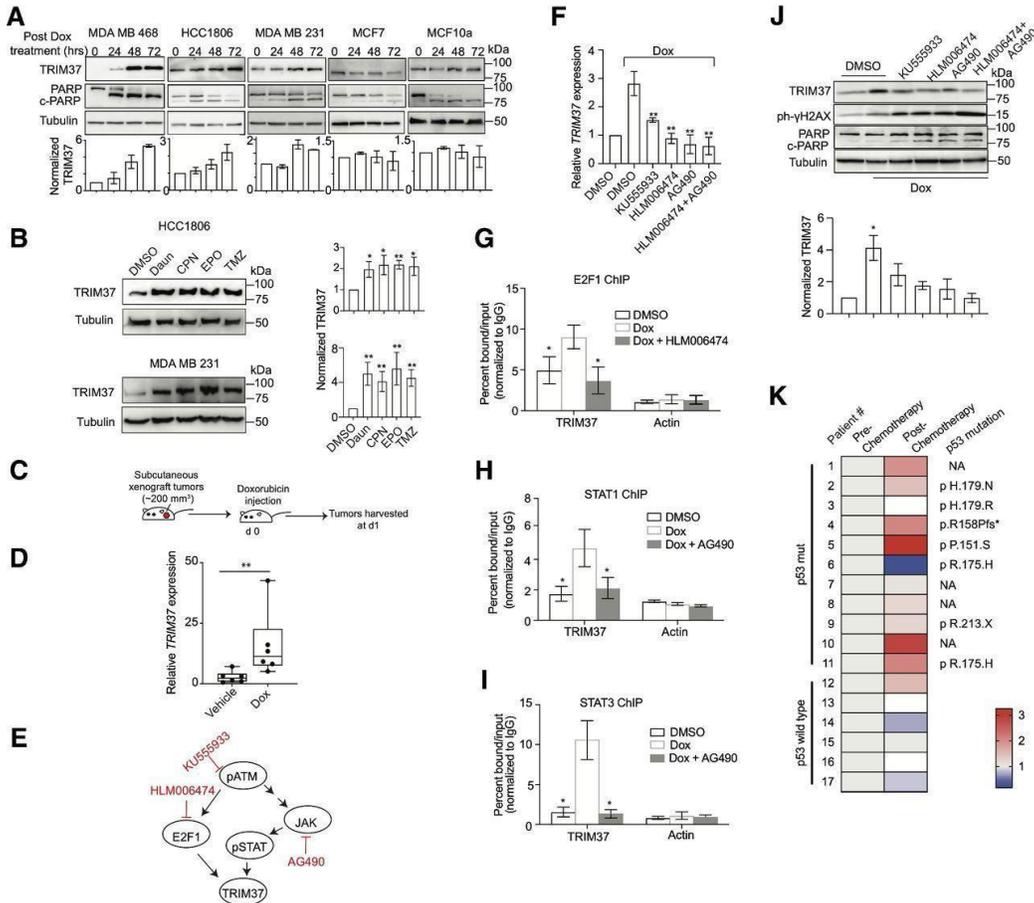
Figure 3-4. **Chemotherapy amplifies oncogenic TRIM37 network in TNBC.** A, Top, immunoblots in MDA-MB-468, HCC1806, MDA-MB-231, MCF6, and MCF10A cells treated with Dox for 0, 24, 48, and 62 hours. Tubulin was the loading control. Bottom, quantification of TRIM37 relative to tubulin. B, Immunoblot monitoring TRIM37 in HCC1806 (top) or MDA-MB-231 (bottom) cells treated with daunorubicin (Daun), cisplatin (CPN), etoposide (EPO), or temozolomide (TMZ). Tubulin was the loading control. Right, quantification of TRIM37 relative to tubulin. C, Schematic showing that NSG mice were treated with i.p. injection of 2 mg/kg Dox once tumor reached the size of approximately 200 mm3. Tumors were harvested postmortem. D, qRT-PCR monitoring TRIM37 expression in HCC1806 subcutaneous tumors derived from NSG mice following treatment with Dox. n = 6 animals per group. E, Schematic of ATM signaling with the downstream effectors of ATM, E2F1, and STAT. Specific small-molecule inhibitors of the ATM signaling are also indicated (red). F, qRT-PCR monitoring TRIM37 expression in MDA-MB-468 cells following treatment with KU555933, HLM006464, or AG490 in combination with Dox. G–I, ChIP analysis monitoring the binding of E2F1 (G), STAT1 (H), and STAT3 (I) to TRIM37 and Actin in MDA-MB-468 cells treated with HLM006464 or AG490 in

combination with Dox. J, Top, immunoblots in MDA-MB-468 cells treated with Dox and the indicated inhibitors of ATM signaling. Tubulin was the loading control. Bottom, quantification of TRIM37 relative to tubulin. K, Heat map for TRIM37 expression in mutant and wild-type TP53 TNBC tumor tissue samples pre- and post-chemotherapy. The type of TP53 mutation is indicated on the right. NA, not available (n = 16). Error bars indicate SD and range of at least three biological replicates. P values were computed using the unpaired t test. *, p < 0.05; **, p < 0.01; ***, p<0.001; ns, not significant.

TRIM37 remodels transcriptional program favoring TNBC metastasis

TRIM37 in association with polycomb complex alters gene expression to promote tumorigenesis [9]. To test whether TRIM37 causes transcriptional misregulation of genes involved in metastasis, we knocked down TRIM37 in MDA-MB-231-D3H2LN-2b [35], hereafter referred to as 231-2b, using TRIM37-specific ASO (TRIM37-ASO) and performed transcriptomic analysis. Of the ~2,600 genes whose expression differed significantly between TRIM37-knockdown and control cells (GSE136616; Fig.3-5A–B), ~61 tumor and metastases suppressors, such as KISS1 and BRMS1, were significantly down-regulated by TRIM37. As robust metastatic suppressors, KISS1 and BRMS1 reciprocally correlate with increased tumor recurrence, metastatic foci and reduced disease-free survival [36-37]. The anti-metastatic function is mediated by altered gene expression through cell signaling pathways [38-39] as well as transcriptional regulation [40-41]. Additionally, gene set enrichment analysis (GSEA) revealed that TRIM37-knockdown downregulates hypoxia, EMT transition, glycolysis, angiogenesis, inflammatory, and immune response-related genes in TNBC cells, indicating TRIM37-dependent activation of a pro-metastatic transcriptional program (Fig.3-

5C). Likewise, KEGG pathway analysis identified TRIM37 target genes that associated with focal adhesion, pathways in cancer, actin cytoskeleton, ECM interaction, and signaling pathways.

To validate the RNA-seq results, we analyzed expression of representative genes in p53−/− MCF10A cells ectopically expressing TRIM37. Expression of KISS1 and BRMS1 was significantly lower in cells ectopically expressing TRIM37 compared with empty vector (Fig.3-5D). Conversely, TRIM37-knockdown tumors expressed KISS1 and BRMS1 at significantly higher levels relative to control xenograft tumors (Fig.3-5E). To investigate the mechanism by which TRIM37 regulate KISS1 and BRMS1, we analyzed binding of polycomb complex components, BMI1 and EZH2, to BRMS1 and KISS1 by directed-ChIP assays. Both the gene promoters were enriched for BMI1 and EZH2 which was diminished after TRIM37 knockdown (Fig.3-5F). These gene promoters were also enriched for H2Aub, which was reduced after TRIM37 knockdown (Fig.3-5F). As expected, knockdown of BMI1 and EZH2 resulted in increased expression of these genes. Our results raised the possibility that TRIM37-mediated repression of metastases suppressors induces transcriptional program favoring metastasis. To test this idea, we analyzed a representative set of 15 genes in the top GSEA categories based on statistical analysis and their known biological functions in multiple steps of metastasis. For all 15 genes analyzed, knockdown of BRMS1 or KISS1 resulted in

their increased expression (Fig.3-5G–H). Consistently, ectopic expression of

TRIM37 in p53−/− MCF10A cells significantly increased expression of all the

TRIM37 target genes compared with empty vector (Fig.3-5I), indicating reciprocal

relationship between TRIM37 and metastases suppressors. To validate RNA-seq

results *in vivo*, a subset of TRIM37 target genes were analyzed in TRIM37-

knockdown tumors. As expected, all the 15 genes analyzed were significantly

reduced in TRIM37-knockdown tumors relative to control tumors (Fig.3-5J).

Moreover, knockdown of TRIM37 also decreased expression of TRIM37 target

genes in MCF6 cells.

To investigate directly the potential function of TRIM37 in metastasis, we

compared the *in vivo* propensity of control and TRIM37-knockdown cells in NSG

mice (Fig.3-5K). Knockdown of TRIM37 showed dramatic reduction in the

metastatic burden in comparison to control tumors that developed in sites

comparable to human breast cancer metastases, such as the brain, lung, liver, lymph

nodes, and bone (Fig.3-5L). TRIM37-knockdown reduced the metastatic tumor

burden in lungs by ~2-fold, bone by ~2-fold, brain by ~3-fold, and liver by ~3-fold

compared to the control animals 21-days after TNBC cell injection (Fig.3-5M).

Further, histological analysis of tumors confirmed the distant tumor growth of the

control and TRIM37-knockdown 231-2b cells in lung and liver. Finally, TRIM37-

knockdown in 231-2b resulted in a modest but significant improvement of post-

injection survival (Fig.3-5N). Collectively, these results demonstrate that TRIM37 overexpression enforces transcriptional program in TNBC tumors that promotes metastatic progression.
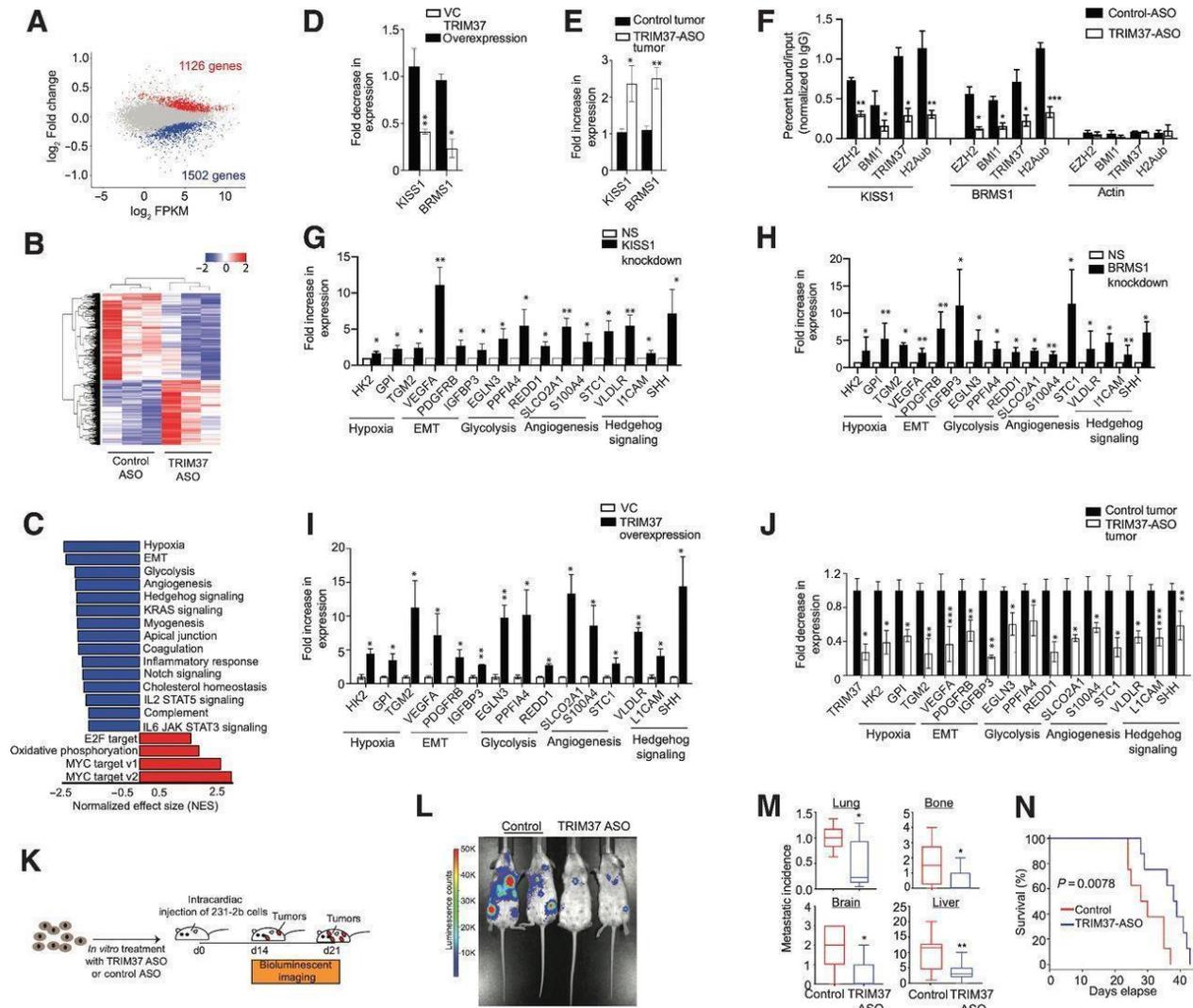
Figure 3-5. TRIM37 alters the transcription program to favor metastatic growth of TNBC tumors. A, MA plot illustrates differential gene expression in TRIM37-ASO–treated compared with control 231–2b cells. Red, significantly upregulated genes (n = 1,126); blue, significantly downregulated genes (n = 1,502); gray, genes not significantly changed (n = 12,440). FDR < 0.05. B, Hierarchical clustering of median-centered gene expression in control or TRIM37-ASO–treated 231–2b cells. Each colored line in the dendrogram identifies a different gene (n = 3). C, Pathways significantly downregulated (blue) or upregulated (red) in TRIM37-ASO–treated cells relative to control 231–2b cells identified by GSEA. D and E, qRT-PCR monitoring TRIM37-regulated metastasis suppressor genes in TRIM37 overexpressing p53−/− MCF10A cells relative to vector control (D) and TRIM37-ASO–treated 231–2b tumors relative to control tumors (E). F,

ChIP monitoring BMI1, EZH2, TRIM37, and H2Aub binding at KISS1, BRMS1, and Actin in control and TRIM37-ASO–treated 231–2b cells. G and H, qRT-PCR monitoring TRIM37 target genes in 231–2b cells expressing KISS1 shRNA (G) and BRMS1 shRNA (H) relative to nonsilencer (NS) shRNA. I and J, qRT-PCR monitoring TRIM37 target genes in TRIM37 overexpressing p53−/− MCF10A cells relative to vector control (I) and TRIM37-ASO–treated 231–2b tumors relative to control tumors (J). K, Schematic showing that mice were injected with control or TRIM37-ASO–treated 231–2b cells intracardially and monitored for metastatic tumor burden. n = 8 animals per group. L, Representative ventral BLI of 231–2b expressing control or TRIM37 knockdown (TRIM37-ASO) at day 21. M, Analysis of metastatic tumor growth in mice tissues measured by relative luciferase signal for lungs and number of metastatic lesions in lungs, bones, brain, and liver. n = 8 animals per group. N, Kaplan–Meier survival curve for mice injected with control or TRIM37-ASO–treated 231–2b cells. n = 8 animals per group. Error bars indicate SD and range of at least three biological replicates. P values were computed using the unpaired t test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant.

Targeting TRIM37 to prevent metastasis in TNBC

Next, we asked whether TRIM37 targeting will reduce metastatic lesions in syngeneic spontaneous metastasis murine model. To this end, Balb/c mice bearing mammary fat pad tumors derived from murine 4T1 cells were administered 1.2 mg/kg of control or smart nanoparticles (Fig.3-6A). To inhibit TRIM37, we utilized smart nanoparticles conjugated with murine cross-reactive anti-FOLR1. While ~90% of the control nanoparticle treated mice developed overt lung metastasis, smart nanoparticles-treated animals showed a dramatic decrease in lung metastasis, with ~60% of smart nanoparticles-treated animals showing no detectable lung metastases (Fig.3-6B). Animals were sacrificed at day 30 post-treatment due to the moribund condition of control nanoparticles-treated mice, which correlated with more rapid tumor growth at the primary and metastatic sites. The lung metastases were confirmed by H&E staining (Fig.3-6C), the gross lung tissue isolated post-mortem (Fig.3-6D–E), and a high proliferative index as determined by Ki67

staining (Fig.3-6F). Furthermore, TRIM37 was significantly reduced in the metastatic lesions isolated from the smart nanoparticle-treated animals compared to control animals. No significant changes in liver histology or serum AST and ALT levels between the control and smart nanoparticles-treated animals indicated a lack of hepatotoxicity.

Primary tumors are routinely treated with a combination of surgery and chemotherapy. Our findings revealed that TRIM37 augments the metastatic potential of TNBC tumors by promoting survival under chemotherapeutic stress, and by inducing metastatic effectors (summarized in Fig.3-6G). These results raised the possibility that combining TRIM37 inhibition with chemotherapy will simultaneously increase chemotherapy efficacy and prevent metastatic progression of TNBC to increase the overall survival in TNBC patients.

To test this idea in a clinically relevant setting, we generated primary tumors by subcutaneously implanting 231-2b lung-tropic cells in female NSG mice (Fig.3-6H). Animals were treated with either control or smart nanoparticles intranasally in combination with a single dose of 2 mg/kg of Dox intraperitoneally post-tumor resection. Significantly, animals treated with control nanoparticles developed lung metastases, whereas smart nanoparticles treatment dramatically reduced metastatic burden in the lungs (Fig.3-6I). H&E-stained lung sections (Fig.3-6J) and luciferase signals from gross lung tissues revealed ~5-fold decrease in metastatic growth in

animals treated with smart nanoparticles compared to control nanoparticle treated animals (Fig.3-6K–L). Tumors from smart nanoparticles and Dox-treated animals showed decreased tumor growth as indicated by significantly higher staining for Caspase 3 (Fig.3-6M) and lower Ki67 staining in lung metastatic tumors in comparison to the control tumors.
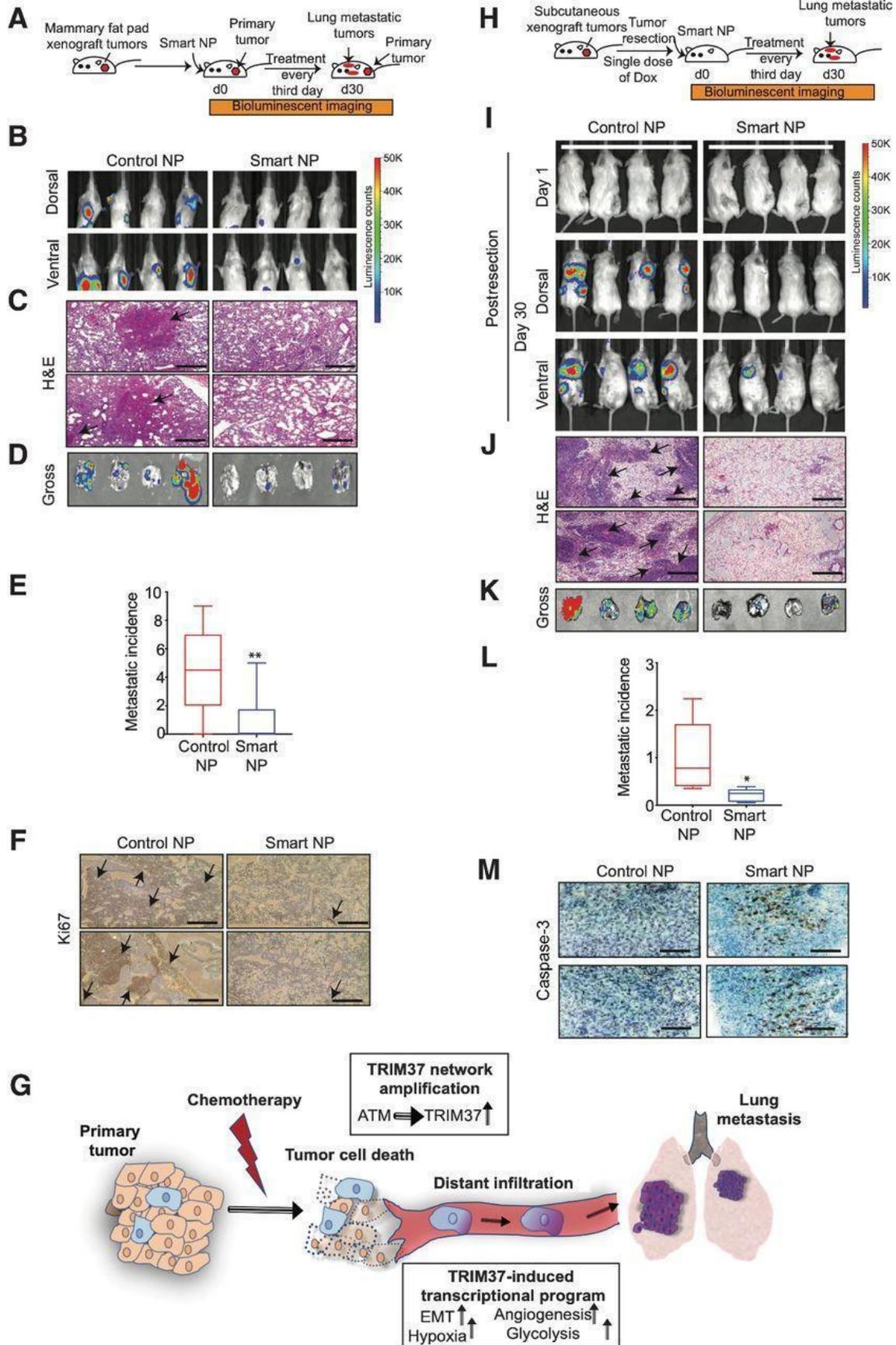
Figure 3-6. Targeting of TRIM37 suppresses metastatic lung tumors *in vivo*. A, Schematic showing that female Balb/c mice bearing mammary fat pad 4T1 tumors were treated with intranasal and intratumor injections of smart or control nanoparticles. n = 8 animals per group. B, Representative dorsal and ventral BLI images of tumor-bearing mice at day 30 after treatment with either control or smart nanoparticles. The color scale depicts the luminescence counts emitted from the metastasis cells. n = 4 animals per group. C, Representative 10× H&E staining images of the lung metastases for control and smart nanoparticle–treated animals. Scale bar, 0.5 mm. Arrows, lung metastatic nodules. D, Lung necropsies from animals in B were analyzed by fluorescent imaging for tumor burden.E, Quantification of metastasis incidence in the lung tissue after control or smart nanoparticles treatment. n = 8 animals per group. F, Ki66 staining of lung tumors derived from mice treated with control or smart nanoparticles. Scale bar, 0.5 mm. Arrows, highly proliferative lung metastatic nodules. G, Model depicting TRIM37 function in multiple steps of TNBC metastasis. H, Schematic showing that NSG mice bearing subcutaneous 231–2b tumors were treated with intranasal injections of smart or control nanoparticles in combination with Dox posttumor resection. n = 8 animals per group. I, Representative BLI images for tumor-bearing mice at day 1 and day 30 postprimary tumor resection. The color scale depicts the luminescence counts emitted from the metastasis cells. n = 4 animals per group. J, Representative 10× H&E staining images of the lung sections of control and smart nanoparticle–treated animals. Scale bar, 0.5 mm. Arrows, lung metastatic nodules. K, Lung necropsies from animals in I were analyzed for tumor burden. L, Quantification of accumulated luciferase signal from the lung tissue after control and smart nanoparticles treatment. n = 8 animals per group. M, Caspase-3 staining of lung tumors derived from mice treated with control or smart nanoparticles in combination with Dox. Scale bar, 0.5 mm. Error bars indicate SD and range of at least three biological replicates. P values were computed using the unpaired t test. *, p < 0.05; **, p < 0.01; ***, p<0.001; ns, not significant.

3.5 Discussion

This study identifies a new TRIM37 network, which is amplified by chemotherapeutic drugs, as a unifying mechanism that drives chemoresistance and metastatic phenotype in TNBC tumors. The results are relevant to ~80% of TNBC patients that lack functional p53 and rely on systemic chemotherapeutic treatments due to the unavailability of any targeted TNBC therapy.

Chemoresistance, in general, is accompanied with extensive genetic and epigenetic alterations. However, whether selection of the clonal cancer cells or new mutations drive chemoresistance phenotype remain to be resolved [32]. A recent genomic and

phenotypic evolution profiling of TNBC tumors identified both pre-existing resistant genotypes as well as transcriptional reprogramming of resistant signatures in TNBC tumors [30]. Our findings are in line with Kim et. al showing that pre-existing higher levels of TRIM37 in TNBC tumors promote resistance to chemotherapeutic stress and thus, increase survival of TNBC cells. On the other hand, chemotherapy triggers ATM signaling to transcriptionally upregulate TRIM37, which could further select for aggressive TNBC cells. In summary, TRIM37-positive TNBC tumors are protected and thrive under continued chemotherapy to cause aggressive metastatic disease.

A major hurdle in finding cure for aggressive TNBC is the lack of known drivers of the metastatic transition, in part due to a lack of mechanistic insights in the development of metastatic tumors. A genomics-driven discovery of recurring genetic mutations and epigenetic aberrations in the breast cancer genome has revealed tumorigenic drivers [6,43]. However, whether these drivers that were primarily identified in the primary tumors are maintained throughout the chemotherapy regimen and the multistep process of metastasis remain to be evaluated. We used genomic and genetic approaches in relevant TNBC cellular, preclinical murine models and tumor biopsies to establish TRIM37 function in reducing therapy-induced DNA damage, increasing cancer cell survival, and causing transcriptional aberrations (Fig.3-7G). Our proof-of-concept results thus

provide a rationale to target such common molecular effectors in combination with chemotherapy to prevent or significantly delay the metastatic progression in TNBC patients.

While targeted therapies are desperately needed to limit damage to healthy tissues, new delivery mechanisms for cancer cell-specific targeting are also required to reduce detrimental side effects in healthy tissues. Molecularly targeted nanoparticles represent one such mechanism and are being aggressively explored for developing new treatment designs. As such, there are four nanoparticle-based therapies in clinical trials - BIND-014 for NSCLC and prostate cancer [44], CALAA-01 for solid tumors [45], SEL-068 for nicotine addiction [46], and Yale BNP for skin cancer [47].

Similar to mechanisms used in antibody-drug conjugates, we utilized a clinically investigative monoclonal antibody, which selectively delivers TRIM37-ASO into TNBC cells. The molecularly targeted nanoparticles offer a significant advantage over antibody-drug conjugates in terms of higher payload concentrations in tumor cells by enhancing retention times and permeability. An important consideration for nanoparticle delivery designs is clearance from the mononuclear phagocytic system. This is particularly critical for therapies designed to treat metastatic TNBC because premature elimination from circulation will prevent uptake by circulating tumor cells, decrease their accumulation in cancer cells and

minimize their therapeutic impact. To overcome these issues, we incorporated PEG into our design, which enables steric stabilization of nanoparticles, and prevents interaction of the nanoparticles with immune cells [48]. Functionalizing nanoparticles with "self" markers or homing molecules can further improve the systemic delivery of nanoparticles [49].

Notably, the majority of single-agent therapies tested to date for metastatic TNBC achieved an unimpressive response rate of less than 20%, with minimal impact on patient survival [50]. The effective and selective delivery of TRIM37-ASO by Farletuzumab-conjugated nanoparticles provides an excellent opportunity to test additional TNBC-enriched surface proteins using monoclonal or bi-specific antibody formats. Some of the targets that can be exploited include TNBC-enriched MUC1, Trop-2 and VEGFR2. Our results also raise the possibility that molecularly targeted nanoparticles can deliver diverse payloads selectively to cancer cells. In conclusion, our results identify a new driver of metastatic progression in TNBC patients and provide a mechanistic link between the two clinically linked phenotypes: chemoresistance and metastasis. Our findings also raise the possibility of clinically targeting TRIM37 to diminish the resistance to therapy, reduce the dissemination of cancer cells, and infiltration of distant sites. We demonstrate that our therapeutic design selectively inhibits TRIM37 and attenuates metastatic progression of TNBC tumors *in vivo*.

3.6 References

1.Pal SK, Childs BH, Pegram M. Triple negative breast cancer: unmet medical needs. Breast Cancer Res Treat 2011;125:627–36

2.Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 2007;109:1721–8

3.Bertucci F, Ng CKY, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, et al. Genomic characterization of metastatic breast cancers. Nature 2019;569:560–4

4.Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell 2017;32:169–84 e7

5.Ryan RJ, Bernstein BE. Molecular biology. Genetic events that shape the cancer epigenome. Science 2012;336:1513–4

6.Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell 2013;153:38–55

7.Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400–4

8.Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, et al. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct Target Ther 2019;4:62.

9.Bhatnagar S, Gazin C, Chamberlain L, Ou J, Zhu X, Tushir JS, et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. Nature 2014;516:116–20

10.Bhatnagar S, Green MR. TRIMming down tumor suppressors in breast cancer. Cell Cycle 2015;14:1345–6

11.Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. Cell 2017;168:670–91

12.Acharyya S, Oskarsson T, Vanharanta S, Malladi S, Kim J, Morris PG, et al. A CXCL1 paracrine network links cancer chemoresistance and metastasis. Cell 2012;150:165–78

13.Hu G, Chong RA, Yang Q, Wei Y, Blanco MA, Li F, et al. MTDH activation by 8q22 genomic gain promotes chemoresistance and metastasis of poor-prognosis breast cancer. Cancer Cell 2009;15:9–20

14.Weiss MB, Vitolo MI, Mohseni M, Rosen DM, Denmeade SR, Park BH, et al. Deletion of p53 in human mammary epithelial cells causes chromosomal instability and altered therapeutic response. Oncogene 2010;29:4715–24

15. Volk-Draper LD, Rajput S, Hall KL, Wilber A, Ran S. Novel model for basaloid triple-negative breast cancer: behavior *in vivo* and response to therapy. Neoplasia 2012;14:926–42

16. Shivange G, Urbanek K, Przanowski P, Perry JSA, Jones J, Haggart R, et al. A Single-Agent Dual-Specificity Targeting of FOLR1 and DR5 as an Effective Strategy for Ovarian Cancer. Cancer Cell 2018;34:331–45 e11

17. Lou S, Zhao Z, Dezort M, Lohneis T, Zhang C. Multifunctional Nanosystem for Targeted and Controlled Delivery of Multiple Chemotherapeutic Agents for the Treatment of Drug-Resistant Breast Cancer. ACS Omega 2018;3:9210–9

18. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun 2016;7:11479.

19. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, et al. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell 2008;29:69–80

20. Shanbhag NM, Rafalska-Metcalf IU, Balane-Bolivar C, Janicki SM, Greenberg RA. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. Cell 2010;141:970–81

21. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature 2012;481:287–94

22. Goldstein M, Kastan MB. The DNA damage response: implications for tumor responses to radiation and chemotherapy. Annu Rev Med 2015;66:129–43

23. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell 2010;40:179–204

24. Williams RS, Williams JS, Tainer JA. Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. Biochem Cell Biol 2007;85:509–20

25. Ismail IH, Andrin C, McDonald D, Hendzel MJ. BMI1-mediated histone ubiquitylation promotes DNA double-strand break repair. J Cell Biol 2010;191:45–60

26. Cohn MA, D'Andrea AD. Chromatin recruitment of DNA repair proteins: lessons from the fanconi anemia and double-strand break repair pathways. Mol Cell 2008;32:306–12

27. Kastenhuber ER, Lowe SW. Putting p53 in Context. Cell 2017;170:1062–78

28. Gasco M, Yulug IG, Crook T. TP53 mutations in familial breast cancer: functional aspects. Hum Mutat 2003;21:301–6

29. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell 2017;32:155–68 e6

30.Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance

Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing.

Cell 2018;173:879–93 e13

31.O'Connor MJ. Targeting the DNA Damage Response in Cancer. Mol Cell

2015;60:547–60

32.Navin NE. Tumor evolution in response to chemotherapy: phenotype versus

genotype. Cell Rep 2014;6:417–9

33.Wu G, Song L, Zhu J, Hu Y, Cao L, Tan Z, et al. An ATM/TRIM37/NEMO

Axis Counteracts Genotoxicity by Activating Nuclear-to-Cytoplasmic NF-kappaB

Signaling. Cancer Res 2018;78:6399–412

34.Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, et al. High-

resolution clonal mapping of multi-organ metastasis in triple negative breast cancer.

Nat Commun 2018;9:5079.

35.Tominaga N, Kosaka N, Ono M, Katsuda T, Yoshioka Y, Tamura K, et al. Brain

metastatic cancer cells release microRNA-181c-containing extracellular vesicles

capable of destructing blood-brain barrier. Nat Commun 2015;6:6716.

36.Ulasov IV, Borovjagin AV, Timashev P, Cristofanili M, Welch DR. KISS1 in

breast cancer progression and autophagy. Cancer Metastasis Rev 2019;38:493–506

37.Kodura MA, Souchelnytskyi S. Breast carcinoma metastasis suppressor gene 1 (BRMS1): update on its role as the suppressor of cancer metastases. Cancer Metastasis Rev 2015;34:611–8

38.Cicek M, Fukuyama R, Welch DR, Sizemore N, Casey G. Breast cancer metastasis suppressor 1 inhibits gene expression by targeting nuclear factor-kappaB activity. Cancer Res 2005;65:3586–95

39.Wu P-Y, Yu I-S, Lin Y-C, Chang Y-T, Chen C-C, Lin K-H, et al. Activation of Aryl Hydrocarbon Receptor by Kynurenine Impairs Progression and Metastasis of Neuroblastoma. Cancer Research 2019:canres.3272.2018 [DOI] [PubMed]

40.Cicek M, Fukuyama R, Cicek MS, Sizemore S, Welch DR, Sizemore N, et al. BRMS1 contributes to the negative regulation of uPA gene expression through recruitment of HDAC1 to the NF-kappaB binding site of the uPA promoter. Clin Exp Metastasis 2009;26:229–37

41.Hurst DR, Xie Y, Vaidya KS, Mehta A, Moore BP, Accavitti-Loper MA, et al. Alterations of BRMS1-ARID4A interaction modify gene expression but still suppress metastasis in human breast cancer cells. J Biol Chem 2008;283:7438–44

42.Ginter PS, McIntire PJ, Cui X, Irshaid L, Liu Y, Chen Z, et al. Folate Receptor Alpha Expression Is Associated With Increased Risk of Recurrence in Triple-negative Breast Cancer. Clin Breast Cancer 2017;17:544–9

43.Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70

44.Sanna V, Pala N, Sechi M. Targeted therapy using nanotechnology: focus on cancer. Int J Nanomedicine 2014;9:467–83

45.Zuckerman JE, Gritli I, Tolcher A, Heidel JD, Lim D, Morgan R, et al. Correlating animal and human phase Ia/Ib clinical data with CALAA-01, a targeted, polymer-based nanoparticle containing siRNA. Proc Natl Acad Sci U S A 2014;111:11449–54

46.Desai RI, Bergman J. Effects of the Nanoparticle-Based Vaccine, SEL-068, on Nicotine Discrimination in Squirrel Monkeys. Neuropsychopharmacology 2015;40:2207–16

47.Deng Y, Ediriwickrema A, Yang F, Lewis J, Girardi M, Saltzman WM. A sunblock based on bioadhesive nanoparticles. Nat Mater 2015;14:1278–85

48.Suk JS, Xu Q, Kim N, Hanes J, Ensign LM. PEGylation as a strategy for improving nanoparticle-based drug and gene delivery. Adv Drug Deliv Rev 2016;99:28–51

49.Alqaraghuli HGJ, Kashanian S, Rafipour R. A Review on Targeting Nanoparticles for Breast Cancer. Curr Pharm Biotechnol 2019;20:1087–107

50.Vidula N, Bardia A. Targeted therapy for metastatic triple negative breast cancer: The next frontier in precision oncology. Oncotarget 2017;8:106167–8

51.Siddharth S, Sharma D. Racial Disparity and Triple-Negative Breast Cancer in African-American Women: A Multifaceted Affair between Obesity, Biology, and Socioeconomic Determinants. Cancers (Basel). 2018;10(12):514. Published 2018 Dec 14.

Chapter 4: The TRIM37 variant rs57141087 contributes to triple-negative breast cancer outcomes in Black women

4.1 Abstract

Triple-negative breast cancer (TNBC) disproportionately affects younger Black women, who show more aggressive phenotypes and poorer outcomes than women of other racial identities. While the impact of socio-environmental inequities within and beyond health systems is well documented, the genetic influence in TNBC-associated racial disparities remains elusive. Here, we report that cancer-free breast tissue from Black women expresses TRIM37 at a significantly higher level relative to White women. A reporter-based screen for regulatory variants identifies a non-coding risk variant rs57141087 in the 5' gene upstream region of the TRIM37 locus with enhancer activity. Mechanistically, rs57141087 increases enhancer–promoter interactions through NRF1, resulting in stronger TRIM37 promoter activity. Phenotypically, high TRIM37 levels drive neoplastic transformations in immortalized breast epithelial cells. Finally, context-dependent TRIM37 expression reveals that early-stage TRIM37 levels affect the initiation and trajectory of breast cancer progression. Together, our results indicate a genotype-informed association of oncogenic TRIM37 with TNBC risk in Black women and implicate TRIM37 as a predictive biomarker to better identify patients at risk of aggressive TNBC.

4.2 Introduction

Breast cancer is the second most common cancer in women, with 297,790 estimated new cases in the United States in 2023 [1]. The relatively high 5-year overall survival of 90.8% indicates that the disease is curable. The advanced stage at the diagnosis and limited access to treatment options contribute to 43,000 annual deaths, but triple-negative breast cancer (TNBC) remains the major cause of mortality in breast cancer patients. TNBCs constitute 12–17% of all breast cancer subtypes, affecting younger women more frequently than patients in other age groups [2]. Histologically, TNBCs are characterized by less than 1% cellular expression of progesterone and estrogen receptors and 0 and 1+ expression of human growth factor receptor 2. Clinically, TNBC tumors are associated with later stage at diagnosis, increased mortality, recurrence, metastatic patterns, and poor prognosis [3]. Given the nonavailability of targeted therapeutic options and the highly invasive nature of tumors, TNBC is usually managed by conventional chemotherapy, which is often accompanied by resistance, relapse, and severe side effects.

TNBC clinical outcomes are shaped by lifestyle, genetics, epidemiologic, and cultural factors; however, the disease etiology might also depend on biogeographical ancestry. Globally, TNBC mortality and prevalence are highest

among women with African ancestry [4-6] who belong to geographically diversified regions, such as Europe, South Africa, and admixed African American populations in the United States [7, 8]. Whether earlier onset or advanced stage at the time of diagnosis, an aggressive tumor phenotype is a characteristic feature of TNBC in Black women (BW) [9-11]. Consequentially, a 5-year survival rate for TNBC in BW is only 14% compared to 36% in women from other racial categories [11]. We argue that genetic drivers may predispose BW to aggressive TNBC. To this point, several groups utilized transcriptomic profiling of tumor tissues that identified racially and ethnically segregating genetic features. A racial stratification of genome-wide association study (GWAS) data for women with African and European ancestry identified a common risk variant at the TERT-CLPTM1L locus, which increased genetic predisposition to TNBC [12]. Martini et al linked racial identity-specific immune response signatures to outcome disparities in tumors targeted with immunotherapies [13]. Another study linked genes, such as ACOX2, CRYBB2, and MUC1, to increased luminal A and basal-like breast cancer risk and survival disparity [14]. A comparative transcriptomic analysis of TNBC cellular models identified differences in GLI1 and Notch1 pathways predisposing African American women to aggressive growth and metastatic TNBC progression [15]. While these studies highlighted inherent genetic and epigenetic differences that vary by racial identity or ethnicity, the functional association of the gene with breast

cancer risk or racial disparity is lacking. Accordingly, these reports underscore the need for mechanistic studies to demonstrate the contribution of genetic factors in breast cancer onset and progression. The outcomes from such studies could inform the methods for early detection, intervention, and prevention of TNBC.

An important benefit of understanding the earliest changes of cancer initiation is the potential to identify biomarkers for breast cancer susceptibility, provide health benefits to the general population, and opportunities for cancer prevention. Over the last decade, greater than 100 cancer predisposition genes conferring a higher risk have been identified through genome-wide mutational analysis, linkage analysis, and candidate gene analysis [16]. For example, the genomic loci harboring well-characterized breast cancer genes, such as BRCA1, TP53, BRIP1, and PTEN, are associated with breast cancer risk [17]. Another longitudinal study identified an aberrant transcriptional regulation of 69 genes in the normal, cancer-free breast tissue of women at high risk of breast cancer [18]. Together, these studies provided critical insight into the genetic and epigenetic changes that predispose to breast cancer and accompany the early stages of carcinogenesis.

Our previous studies comprehensively described tripartite motif-containing protein 37 (TRIM37) function as a novel breast cancer oncoprotein [19, 20]. TRIM37 is highly expressed in breast cancer tissue and is associated with poor overall patient survival [19]. TRIM37 represses gene transcription by mono-ubiquitinating histone

H2A at Lys119 (H2Aub) at various genomic loci, including tumor suppressors and pro-apoptotic genes [19]. TRIM37 overexpression renders RAS-transformed, premalignant breast cells tumorigenic, and inhibition of TRIM37 function reduces tumor growth in xenograft and patient-derived mouse models [19]. We also recently demonstrated that TRIM37 drives aggressive TNBC biology through a transcriptional program favoring resistance to chemotherapy and metastasis [21]. However, in contrast to established tumors, the high-resolution functional characterization of TRIM37 during the initial steps of neoplastic transformations before disease manifests is still limited.

Here, we describe the mechanism by which TRIM37 associates with breast cancer risk and aggressive phenotype in BW. We show that TRIM37 upregulation in BW breast epithelial cellular systems triggers transcriptional reprogramming, favoring loss of identity and acquisition of stemness and mesenchymal phenotypes. Using comprehensive genomic and functional analysis, we identify an ancestry-specific, non-coding regulatory variant rs57141087 predominant in BW that modulates TRIM37 levels by increasing enhancer–promoter contact via NRF1. We show that TRIM37 overexpression in the early stages confers neoplastic transformations, accelerates tumorigenesis, and drives cells into a fate of malignancy. Thus, we provide functional proof for the effects of TRIM37 in TNBC etiology that are determined by the SNP rs57141087 in BW.

4.3 Materials and Methods

Cell lines and reagents

MDA MB 231, MCF10A, and HEK293T cells were cultured as previously described [19]. KTB51, KTB37, and KTB39 cells (Harikrishna Nakshatri, Indiana University) were maintained in a mixture of DMEM/F-12 and DMEM (3:1) supplemented with 5% fetal bovine serum (FBS, Avantor), insulin (5 μg/ml, Gibco), EGF (20 ng/ml, Invitrogen), hydrocortisone (0.4 μg/ml, Invitrogen), adenine phosphate salt (24 μg/ml, Invitrogen), and Rock inhibitor (5 μg/ml, Invitrogen) [23]. All cells were kept under the same conditions at 37 °C and 5% CO_2. Cells cultured at the same time were pooled together and seeded after counting in a 6-well or 10-cm dish. Following this, cells were subjected, in a random order, to treatment with shRNA, sgRNA, plasmids, or the control. Cells were routinely tested for mycoplasma using the PlasmoTest kit (Invitrogen) [21].

Animal care

NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (Jackson Laboratory) mice were housed in a specialized pathogen-free facility accredited by the American Association of Laboratory Animal Care [21]. All animal studies and protocols (#22405 and #22406) were approved by the guidance of the University of California, Davis Animal Care, and Use Committee (IACUC).

TRIM37 and RAS overexpression

To overexpress TRIM37 and RAS, cells were seeded in a six-well plate to achieve 60–80% confluency and subsequently transduced with 200–500 µl lentiviral particles carrying the TRIM37 Lenti ORF clone, mGFP-tagged (Origene), pBabe Rasmut (Addgene), or both in 2 ml of suitable medium supplemented with ~6–10 µg/ml polybrene. For selection, the media was replaced with puromycin (1 µg/ml), and stable cells were maintained.

Doxycycline-mediated activation of TRIM37

A TRIM37 cDNA clone (Origene) was subcloned into the vector pTRE-2 using BamH1 sites, and insertion was confirmed by Sanger sequencing. The TRIM37-expressing vector and pTET-ON (Addgene) were co-transfected into cells using Effectene reagent (Qiagen). After 24 h, the media containing the appropriate selection agents were replaced. Stable cells were induced with 500 ng/mL doxycycline (Dox) at various time points.

RNA interference

To achieve stable shRNA knockdowns, cells were seeded as mentioned above and transduced with 200–500 µl of lentiviral particles expressing shRNAs (obtained from Open Biosystems/Thermo Scientific through the UMMS RNAi Core Facility, listed in the reagents and tools table) or sgRNA in a total volume of 2 ml of appropriate media supplemented with ~6–10 µg/ml polybrene. Media was replaced

after overnight incubation, and cells were subjected to puromycin selection

(1 µg/ml) for at least 3 days, as described previously [21].

Immunoblotting

Cell lysates were prepared by lysing the cell pellet in RIPA buffer

supplemented with one mM Dithiothreitol and 1× protease inhibitor. Immunoblots

were probed using antibodies against Gapdh (Thermofisher), TRIM37 (Cell

Signaling), and RAS (Cell Signaling).

Quantitative RT-PCR (qRT-PCR)

Total RNA was extracted, and qRT-PCR was performed as previously

described [22], with gene expression normalized to GAPDH. Primers used are listed

in the reagents and tools table. Each sample was analyzed at least three independent

times, and results from at least three different biological triplicates are presented.

Proliferation assay

Briefly, cells were seeded at a density of 2×104 cells/ml in a 12-well plate.

Cells were trypsinized and resuspended in 1 ml of complete media. Ten microliters

of the resuspended cells were mixed with 10 µl of trypan blue (1:1). The number of

viable cells was determined after 48, 96, 120, and 144 h using the Countess 3 Cell

Counter (Invitrogen).

Wound-healing assay

The cell migration was evaluated by seeding cells in a 48-well plate at a density of 2×10^6 cells/ml. After 24 h of incubation, culture inserts were removed using sterile pipetting tips, and the cell monolayer was gently washed with PBS to remove any cell debris and unattached cells. Fresh medium was then added to the wells. Plates were photographed immediately, and cell migration was monitored and recorded over a period of 4 h.

Colony formation assay

Cells were seeded in six-well plates and cultured until visible colonies were formed. Colonies were fixed with 95% ethanol and 5% glacial acetic acid at 37 °C. After 30 min, colonies were washed twice with PBS and stained with 0.5% Crystal Violet dissolved in 2% ethanol and 98% PBS. Cells were imaged using a ChemiDoc Imaging System (Bio-Rad), and colonies were counted.

Mammosphere assay

Five hundred cells were seeded in ultra-low-adherence 24-well plates (Corning) and cultured for 10 days. Images were captured with a ×10 microscope, and mammospheres were counted. For qRT-PCR and flow cytometry, mammospheres were dissociated with 1× trypsin and resuspended in the appropriate media.

Flow cytometry

For CD24/CD44 staining, dissociated mammospheres were washed
sequentially with PBS and FACS buffer and then resuspended in 100 µl of FACS
buffer. Fluorescent-conjugated antibodies, including PE-conjugated anti-CD24 and
Pacific, Blue-conjugated anti-CD44 antibodies (BD Biosciences, San Jose, CA,
USA), were added. Cells were incubated for 45 min at 4 °C. Flow cytometry
analysis was performed using a flow cytometer, and FlowJo software was used for
data analysis.

For EpCAM staining, cells were incubated with a non-conjugated primary
antibody against EpCAM (BD Biosciences, San Jose, CA, USA) for 45 min at 4 °C.
After washing three times, the cells were incubated with a PE-conjugated secondary
antibody for an additional 45 min at 4 °C. Flow cytometry analysis was performed
as described above.

NSG tumor xenograft with 17β-estradiol tablet

Age-matched female NSG mice were subcutaneously injected with one
estradiol tablet (0.18 mg/pellet, Fisher Scientific) and 10 million of the indicated
cell lines. For the doxycycline-inducible TRIM37 overexpression experiments,
NSG mice were fed with a doxycycline diet (ENVIGO, 625 Doxycycline).

Ki67, hematoxylin, and eosin staining

Tumors were excised and then fixed in 4% paraformaldehyde (PFA). Tumors
were sectioned and stained with Ki67, H&E, and antibodies against ER, PR, and

Her2. The in situ hybridization (ISH) Scoring Algorithm was used to score ER, PR, and Her2 [24].

Vector construction

A 960-bp fragment (chr17: 59,106,446–59,107,405, hg38) of TRIM37 5'UTR was amplified from MDA MB 231 and cloned into the KpnI and NheI sites of pGL3-Basic (Promega) vector. In addition, 353-bp (chr17: 59,107,053–59,107,405, hg38) and 909-bp (chr17: 59,107,053–59,107,961, hg38) fragments of TRIM37 5'UTR were also amplified from MDA MB 231 and inserted into the KpnI and NheI sites of the pGL3-promoter (Promega) vector. Risk alleles of individual SNPs were introduced using the Q5 site-directed mutagenesis kit (New England Biolabs), with all constructs confirmed for correct variant incorporation by Sanger sequencing.

Transfection and luciferase assay

Cells were plated 1 day before transfection for luciferase assay. Cells were transiently transfected with pBABE-puro (Addgene) combined with pGL3-basic (Promega) or pGL3-basic containing the 960-bp of TRIM37 5'UTR, harboring the reference or risk allele for each indicated SNP. For enhancer activity assay, cells were transfected with pBABE-puro (Addgene) along with pGL3 promoter (Promega) or pGL3-promoter vector containing a 353-bp or 909-bp of TRIM37 5'UTR harboring the reference or risk allele for rs57141087 SNP. The transfection

was performed using Effectene reagent (Qiagen) according to the manufacturer's protocols. Media was replaced after 48 h of incubation, and cells were subjected to puromycin selection (1 µg/ml) until control cells were dead. Cells were then harvested, and luciferase assay (Luciferase Assay System, Promega) was performed according to the manufacturer's instructions.

Single-stranded oligodeoxynucleotides (ssODNs) mediated knock-in using CRISPR/Cas9

sgRNAs targeting sequences adjacent to the rs57141087 region were designed using TargetFinder and cloned into the LentiCRISPRv2 vector. KTB51 cells were transduced with lentiCRISPRv2-sgRNA virus for 24 h, followed by transfection with ssODN (Integrated DNA Technologies). Subsequently, all clones were validated using Next Generation Sequencing Amplicon-EZ and risk SNP-specific restriction digestion with the HhaI enzyme.

Directed-ChIP assay

ChIP assays were performed as described previously [21]. Antibodies against POLII (Abcam), NRF1 (Cell Signaling), and IgG (as a negative control) were used. Primer sequences for amplifying ChIP products are listed in the reagents and tools table. Each ChIP experiment was carried out at least three times independently.

Nuclear extract preparation

Nuclear extract was prepared from either normal HEK293T cells or HEK293T NRF1 knockdown cells. The cell pellet was swollen in 200 μl of ice-cold NAR A buffer for 10 min. Centrifugation was performed to collect the nuclear pellet at 2000 rpm for 2 min at room temperature. Subsequently, the nuclear pellet was lysed in an ice-cold NAR C buffer with vigorous shaking for 25 min. Protein concentration was determined using Pierce's BCA protein assay kit (Thermo Fisher Scientific). The nuclear extracts were stored at −80 °C until further use.

Looping assay

The looping assay was performed, and qRT-PCR results were analyzed as described previously [25]. The beads were digested with XcmI (New England Biolabs) at 37 °C for 5 h.

Chromatin conformation capture (3 C) with qRT-PCR analysis

3 C analysis was performed according to previously established methods [26]. Following preparation of cell nuclei and formaldehyde cross-linking, samples were digested with ApaI (New England Biolabs) overnight at 37 °C. Ligation products were quantified using qRT-PCR assay. To determine primer efficiencies, a serial dilution of BAC clone (BACPAC Resources Center at BACPAC Genomics, Inc.) was digested with ApaI and ligated with T4 DNA Ligase. Primer sequences are listed in the reagents and tools table.

Electrophoretic mobility shift assay (EMSA)

Consensus NRF1 DNA binding sites in the TRIM37 gene locus were identified using the UCSC Genome Browser and the Predicting Regulatory Functional Effect by Approximate P-value Estimation (PERFECTOS-APE) software. NRF1 binding site was prepared by constructing complementary oligonucleotides with the candidate binding site—GCGCGG—in the middle. A mutant version of this site was prepared by exchanging the conserved C and T in each NRF1 consensus binding site, and the primers used are listed in the reagents and tools table. EMSA was performed as described previously [27]. Biotin-labeled probes were detected using the LightShift Chemiluminescent EMSA kit (Thermo Scientific) according to the manufacturer's instructions.

Data analysis

Data source

Transcriptomic data for TNBC were downloaded from The Cancer Genome Atlas (TCGA), Yale University (PRJNA704957), and the University of Alabama (GSE142731). Transcriptomic data from women at high risk of developing breast cancer were retrieved from GEO under the accession number GSE164641. RNA-seq data for normal breast samples was downloaded from TCGA, and genomic data were used for admixture analysis. Transcriptomic data for cancer-free breast tissue were downloaded from TCGA, GTEx, and GEO accession numbers (GSE164641, GSE111601, and GSE277029). Survival analysis data for TNBC patients were

extracted from TCGA, GSEA39004, and GSE18229. Genomic and transcriptomic

datasets for the Yoruba and UTAH populations (Accession: E-GEUV-1) from the

1000 Genomes were used for the eQTL analysis.

SEER data analysis

TNBC incidence and mortality rate statistics were obtained from the

Surveillance, Epidemiology, and End Results database (Surveillance,

Epidemiology, and End Results Program, cancer.gov). GraphPad Prism was used to

create graphical representations of trends in TNBC incidence, survival, and

mortality rates from 2010 to 2020, stratified by age and racial identity. P value

significance was calculated using the proportion test.

TRIM37 expression analysis

Multiple TNBC mRNA datasets were downloaded from TCGA, GSE142731,

and PRJNA704957. GSE164641 provided mRNA data for normal breast samples

from women with an average risk or high risk of developing breast cancer. In

addition, mRNA datasets from normal breast tissue were downloaded from GTEx

(phs000424.v9.p2), TCGA, GSE164641, GSE111601, and GSE277029. These

datasets were merged into a single meta-dataset using a cross-platform meta-

analysis approach, as previously described [28]. Boxplots were used to show the

TRIM37 expression difference between Black and White or average-risk and high-

risk women. Statistical analysis was performed using the t test. All analyses were carried out using R-studio packages.

Multivariate/univariate analyses

To investigate the relationship between TRIM37 and clinicopathological factors (racial identity, age, BMI, cancer risk, menopausal state, parity, and tumor stages), we employed a multivariate or univariate linear regression approach using the R regression function lm.

Chi-square statistics

The chi-square test was used to evaluate the association between TRIM37 expression and clinicopathological variables. The test was run using GraphPad Prism.

Survival analysis

The association between high- or low TRIM37 expression and overall survival in Black and White TNBC patients was examined using survival data from TCGA, GSE39004, and GSE18229. The receiver operating characteristic (ROC) plot was used to determine the optimal cutoff value for dividing TNBC patients into "high" or "low" TRIM37-expressing groups. Cox proportional hazards regression was used to perform survival analysis, and the results were visualized using Kaplan–Meier survival plots.

Admixture analysis

Ancestry estimation was determined using the Admixture v1.3.0 software [29]. Admixture was run using the imputed TCGA breast SNP variants and RNA-seq data from three estimated populations (Black, White, and Asian). The correspondence between the columns and the population was identified using the originally reported racial identity information from three estimated populations (Black, White, and Asian). The analysis results in a proportional breakdown of each sample into three super-populations (African, White, and Asian).

Differential expression gene

RNA was isolated from cancer-free paraffin-archived breast tissue samples of parous and premenopausal Black (n = 3) and White (n = 3) (provided by Natascia Marino) with the AllPrep DNA/RNA FFPE Kit (Qiagen). RNA-seq was performed by Novogene Genome Sequencing Company using the Illumina Novoseq platform with a paired-end 150 bp sequencing strategy. RNA-Seq data was aligned to the human genome assembly GRCm38/hg38 using the HISAT2 software package. Gene quantification and count data were generated using Featurecounts. DEG analysis between the three Black and three White samples was conducted using R-packages Deseq2 and edgeR. Differentially expressed genes were overlaid with TRIM37-TS genes [21]. Gene set enrichment analysis to identify pathways enriched in TRIM37-TS upregulated in Black women was performed through the GSEA 4.3.2 software. Additionally, the association between EMT and CSC hallmark

genes and recurrence-free survival (RFS) for patients with TNBC was examined using the Kaplan–Meier plotter [30]. EMT and CSC signature genes were divided into high and low based on the upper quartile.

Functional annotation of SNPs

Genotype and RNA-seq data from the Genotype-Tissue Expression (GTEx, phs000424.v9.p2) project were used to calculate the normalized effect size (NES) for the association between TRIM37 expression and the expression quantitative trait loci (eQTL) status of the SNPs. Linkage disequilibrium (LD) was calculated between all SNPs (Chr17:59,108,921–58,982,138 Mb) in the African population based on the 1000 Genomes Project data using the LDmatrix Tool [31]. LD-SNPs were defined based on a r2 value greater than 0.8. In addition, mapped read files of epigenetic enhancer marks, notably H3K4me1, H3K27ac, DHS, and H3K4me3 of breast cell lines, were accessed through the Cistrome database. All peaks were identified via WashU Epigenome Browser hg38, and rs57141087 location overlapped with epigenetic enhancer marks.

Transcriptional factors binding site prediction

The TF binding sites for the region harboring rs57141087 were predicted using the UCSC Genome Browser and the Predicting Regulatory Functional Effect by Approximate P value Estimation (PERFECTOS-APE) software, in conjunction

with the transcription factor binding sites (TFBS) motif collection from JASPAR

Parameters for prediction included a P value < 0.05 and a fold change >2.

RNA-seq analysis

RNA was isolated from control and TRIM37 overexpressing cells; Control

and KTB51-G clone with RNeasy Mini Kit (Qiagen). RNA was also isolated from

Formalin-Fixed Paraffin-Embedded normal breast tissue (provided by Natascia

Marino) with AllPrep DNA/RNA FFPE Kit (Qiagen). RNA-seq was performed by

Novogene Genome Sequencing Company using the Illumina Novoseq platform

with a paired-end 150 bp sequencing strategy. RNA-Seq data was aligned to the

human genome assembly GRCm38/hg38 using the HISAT2 software package.

Using the DESeq2 R package, the differential expression of genes in the control and

experimental group was determined. GSEA 4.3.2 software was used for gene set

enrichment analysis for all genes. The volcano plot and heatmap were drawn using

R.

Imputation analysis

Ungenotyped variants were imputed using Beagle v5.2 for TCGA SNP array

data. Beagle is based on identity by descent (IBD) and uses a hidden Markov model

to account for uncertainty in inferring IBD [32]. The imputation used the 1000

Genomes biallelic SNV called GRCh38, released by EMBL-EBI in December

2018, as the reference panel [33]. HapMap genetic maps in PLINK format for GRCh38 were input to Beagle to estimate the recombination rate.

Breast cancer risk analysis

TNBC risk association with the risk allele of rs57141087 was calculated using Genome-wide association studies (GWAS) data for African ancestry [34]. Summary-level statistics data for TNBC (GCST90296722) were downloaded from the GWAS Catalog. The odds ratio (OR) was calculated from the beta coefficient using the following equation: beta = ln(OR).

Statistical analysis

All experiments were performed at least in triplicate, and the results presented are the mean of at least three different biological replicates. The comparisons between the two groups were made by unpaired t test; comparisons between multiple treatment groups were made by one-way or two-way ANOVA with indicated multiple comparisons post hoc tests. The overall survival of high- or low TRIM37 was calculated using the log-rank test. The difference in incidence and mortality rates between Black and White TNBC patients was assessed using the two-proportions z-test.

Data availability

All transcriptomic data generated is available through the Gene Expression Omnibus (GEO) under the accession number GSE277029. The image source data

can be accessed through BioImage Archive submission can be accessed with

accession number S-BIAD1356:

Table 4-1: List of primers and shRNA used in this study

SNPs Mutagenesis	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
rs150880035	GTTTCTCAAAaTGCGGGCCAC	CCTGGTGCAGAAGGTTCATG
rs57141087	TTTCTCAAAGcGCGGGCCACA	CCCTGGTGCAGAAGGTTC
rs1029035382	ACCTGCTGCCaAATCACACGATG	GGTCTGTGGCCCGCACTT
rs914034546	TCACACGATGtTGGGTGAAAACG	TTCGGCAGCAGGTGGTCT
rs1359804461	CACACGATGCgGGGTGAAAAC	ATTCGGCAGCAGGTGGTC
rs904164828	GCCCCAAGCTaAGGCCGCCTG	TCTGCTACGCCACCCACG
rs572032837	CCCCAAGCTCcGGCCGCCTGC	CTCTGCTACGCCACCCACG
rs568274755	GTCTCGTATGtGCCCCGCAAC	TTGGCGACTCGCTGCCTC
rs1042165062	CGCTCCTCCCgTCCTTTCCCC	TGGAGCGATCGCCTAGTTC
Mutagenesis		
NRF1	GTTCCATAGGTTATTTCCCGCGTCACGTGG	TGACGGTGGAGTTCAGCG
NRF1	GCGTCCTGCGATTTTTCGCTTCGCAGC	TGTGGGAGGGGAAAGGAG
NRF1	GCTCAGGCCGTTATTTCAGGCTAGGGCCG	TTGGGGCTCTGCTACGCG
qRT-PCR		
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
TRIM37	AACAGAGCGTGGAGAGCATT	CTTCTGCCCAACGACAATTT
NRF1	TCAGCAAACGCAAACACAGG	AAGTGAGACAGTGCCATCAGG
PAX5	ATGGATGCACACGGATGTTC	TCAACGGCTTGTGTCCTTTG
GFI1	AAAGCAAGAAGGCTCACAGC	GCATTTGAAGTGCTGTCTGC
CDH1(E-cadherin)	ACGCATTGCCACATACACTC	TCGGGCTTGTTGTCATTCTG
CDH2(N-cadherin)	TGTCCTTACTGTTGCTGCAG	TGTAACAGACACGGTTGCAG
SNAIL	TCCAGAGTTTACCTTCCAGCAG	ACAGAGTCCCAGATGAGCATTG
TWIST	TGGCCTGCAAAACCATAGTC	TGCATTTTACCATGGGTCCTC
ZEB1	ATAGCAAGGAGTGGAGCATAGG	AAAGCCACATCAGCAACAGC
Looping_enh	AGAAATGACACCTCAGTTTG	GTGGTTCTCTCTGGAAGGAA
Looping_prom	CCCCTGACGTGGGCGCCGGG	GCCGCTGGCGACCCGCAGGC
ChIP		
TRIM37	AATGAGAAATGACACCTCAGTTTGG	GGTTTGTAGGACTTGCGTTCC
TRIM37	ACTTAACTACGGGTGTGGCTC	CTGGCGTACTGGTGGAAAGC
TRIM37	GCGCGCGCCCGCGTCACGTGGGC	TCCTTTCTCCCGGCTCAG
TRIM37_ex5	AATTGTCGTTGGGCAGAAGA	ACAAAGGCAAAATCCATTCAATATTGA

TRIM37_enh	ACCATGAACCTTCTGCACCA	GAAAGCGCAGACTTCAGAGC
TRIM37_MP	CAGCGTCCTGCGCGCATGCGCTT	GCCCACGTGACGCGGGCGCGCGC
TRIM37_MP	CCTCCTTTCCCCTCCCACA	GAAGAAGGTGCCGCAGAGAA
3C assay		
A	AAGAATGTTTATGTACATTTC	
B	GTGGTACGCTCTCAGCTCACT	
C	CATGCCCAGCTAGATATATAGTCAT	
D	TAATTCTTTTAAGTAAAATT	
E	AATGAGAAATGACACCTCAGTTTGG	
F	TCTGCACCAGGGTTTCTCAAA	
G	ACCTGCTGCCGAATCACACGAT	
H	GCTGGGTGAAAACGCGGATGG	
E	CGTAGCAGAGCCCCAAGCTCA	
J	AGTCGCCAAGTCTCGTATGC	
K	GCCCCGCCGAGAGCCGGAGGC	
Anchor	CTCTGGTTACGGTTGGTGTTT	
Cloning		
TRIM37	GGTACCAATGAGAAATGACACCTCAGTTTGG	GCTAGCCGCTCTGTTCATCCATTGCC
TRIM37	CGGGGTACCTAATTCTTTTAAGTAAAATT	CCCGCTAGCCCACCCACGCCCACC
TRIM37	GGTACCAATGAGAAATGACACCTCAGTTTGG	CCCGCTAGCCCACCCACGCCCACC
EMSA	GGTACCAATGAGAAATGACACCTCAGTTTGG	ATTCGGCAGCAGGTGGTC
sgRNA	CACCGGCATCGTGTGATTCGGCAGC	AAACGCTGCCGAATCACACGATGCC
shRNA	**Catalogue #**	
NRF1_shRNA#1	TRCN0000016903	
NRF1_shRNA#2	TRCN0000016905	
GFI1_shRNA#1	TRCN0000020464	
PAX5_shRNA#1	TRCN0000016060	
PAX5_shRNA#2	TRCN0000016062	

4.4 Results

African ancestry influences TRIM37 expression in TNBC and cancer-free

breast tissue

Almost 30% of all breast cancer cases in BW are predominantly TNBC [11].

We compared the TNBC incidence and survival rates in BW and White women

(WW), excluding Hispanics, by interrogating the Surveillance, Epidemiology, and End Results Program (SEER) for the years 2010 through 2020. The longitudinal analysis over ten years revealed that approximately two-fold higher number of BW was diagnosed with TNBC than WW (Fig. 4-1A). Similar analysis across different age groups (15–39, 40–74, 75+ ages) showed significantly higher TNBC incidence (Fig. 4-1B) and mortality rate (Fig. 4-1C) in BW than WW over ten years (2010–2020). These results sharply contrast the observed decrease in breast cancer incidence in recent years (Supplementary Fig.4-1A), indicating that BW remains susceptible to aggressive TNBC phenotype despite improved treatment and access to health care. Previous studies have identified intrinsic molecular differences in TNBC tumors accounting for aggressive progression in BW compared to WW [15]. We, therefore, queried TRIM37 expression in different stages of TNBC stratified by self-reported BW and WW using datasets available through GSE142731 (Data ref: [35]), PRJNA704957 (Data ref: NCBI Sequence Read Archive PRJNA704957, 2021), and TCGA (Data ref: The Cancer Genome Atlas Program (TCGA-Breast)). A total of 319 patients were included in the analysis. Interestingly, the meta-analysis revealed ~1.63-fold higher TRIM37 expression in early histological Stage I TNBC tumors from BW than in WW (Fig. 4-1D), which was not the case for Stage II–IV (Fig. 4-1E). Univariate analysis confirmed the association between TRIM37 expression in the Stage I TNBC tumors with racial identity ($P = 0.0174$, Fig.

Supplementary Fig.4-1B). Next, we assessed to what extent the early-stage

differences in TRIM37 expression could explain the disparity in the overall survival

of TNBC patients. Indeed, BW with TNBC tumors expressing high TRIM37

showed poor overall survival, with a median survival of ~114 months, relative to

WW, with a median survival of ~245 months (P = 0.0011; Fig. 4-1F). Notably, no

significant differences in overall survival were observed for low TRIM37-

expressing TNBC tumors from BW and WW (Supplementary Fig.4-1C).

Next, to determine whether there is an independent contribution of TRIM37

to the neoplastic transformation in BW, we interrogated TRIM37 expression in

normal, cancer-free breast tissue. Out of 549 samples analyzed, 119 were BW, 430

were WW, 115 were below or equal to the age of 35, and 434 were above the age of

35. Surprisingly, the meta-analysis revealed significantly higher TRIM37 levels in

the breast tissue of BW relative to WW (Fig. 4-1G). No difference in TRIM37

levels was observed when women were stratified by age (Supplementary Fig.4-1D).

Using the $\chi2$ test, we find that higher TRIM37 levels in the breast tissue were

associated with BW (Fig. 4-1H). Univariate analysis demonstrated a significant

association between TRIM37 expression and racial identity but not in age (Fig. 4-

1H). Further, in the multivariate linear regression test, relationships were examined

between the TRIM37 expression on one hand and the racial identity and age on the

other. The test demonstrated that higher TRIM37 levels were more likely to be

associated with BW than WW (P = 0.00233, Fig. 4-1H). Other factors, such as age, did not show a significant association with TRIM37 levels (Fig. 4-1H).

We note that the genomic complexities of admixed populations may obscure the race-specific molecular differences in self-reported ethnicity. We, therefore, estimated the genomic ancestry for each sample in the TCGA cohort, which indicated greater than 80% African ancestry in the six self-reported BW samples and greater than 70% European ancestry in 103 of self-reported WW (Supplementary Fig.4-1E). To investigate the ancestry-specific TRIM37 expression, we isolated analyses to samples with greater than 70% African or European ancestry. We consistently find that the breast tissue of BW expresses significantly higher TRIM37 than WW (Fig. 4-1I). The univariate analysis confirmed a significant association of TRIM37 with ancestry (P = 0.050). These results revealed African ancestry-specific, distinct TRIM37 expression in normal, cancer-free breast tissue.

Given that TRIM37 is associated with genomic instability, tumor progression, and poor prognosis [19, 21], we hypothesized that TRIM37 levels in breast tissue could likely inform the clinical outcome. We, therefore, examined TRIM37 expression in cancer-free breast tissue from women at higher lifetime risk of developing breast cancer. Of 178 healthy, cancer-free breast tissue from women of median age ~40 years, 72 were grouped into a higher lifetime risk of developing

breast cancer based on the Tyrer-Cuzik model (score ≥20%) [18]. Out of 178

samples, 114 were premenopausal, 64 were post-menopausal women, 135 had

carried at least one pregnancy, and 68 had a BMI in a healthy weight range. Thirty

women were Black, and 148 identified as White. Notably, TRIM37 was expressed

significantly higher in the breast tissue of women allocated to a high-risk group

(n = 72) than in the average-risk group (n = 106, Fig. 4-1J). Consistent with

previous results, we find higher TRIM37 levels in BW relative to WW

(Supplementary Fig.4-1F). No significant differences in TRIM37 levels were

identified when samples were stratified by age, BMI, parity, and menstrual status

(Supplementary Fig.4-1G–J).

Next, using the χ2 test, we asked whether TRIM37 expression is associated

with age, menopause status, breast cancer risk, racial identity, parity status, and

BMI. The tests proved that breast cancer risk was significantly higher in samples

with high TRIM37 expression (Fig. 4-1K). To better understand the features of

young women's breast tissue predictive of clinical outcome, the univariate and

multivariate analyses combining clinicopathologic variables with TRIM37

expression value were performed. The univariate analysis confirmed the significant

association between TRIM37 and breast cancer risk (P = 0.00217) and obesity

(P = 0.0355, Fig. 4-1K). We noted near-significant associations of TRIM37

expression with parity (P = 0.0565) and racial identity (P = 0.09851), in part due to a

smaller sample size (Fig. 4-1K). In multivariate analyses, TRIM37 correlated

significantly with breast cancer risk (P = 0.0005), BMI (P = 0.003), and near

significantly with racial identity (P = 0.0568; Fig. 4-1K). These results raised a

possibility that early-stage high TRIM37 observed in the breast tissue of BW could

give cancer cells a "head start," impacting the disease trajectory and outcome.

H

Characteristics	No.(%)	p-value
χ² test		
Racial identity		
White	430(78.32)	0.0097
Black	119(21.68)	
Age		
≥ 35	115(20.95)	0.7622
<35	434(79.05)	
Univariate analysis		
Racial identity		
White	430(78.32)	0.0023
Black	119(21.68)	
Age		
≤35	115(20.95)	0.946
> 35	434(79.05)	
Multivariate analysis		
Racial identity		
White	430(78.32)	0.0023
Black	119(21.68)	
Age		
≤35	115(20.95)	0.8910
> 35	434(79.05)	

K

Charateristics	No. (%)	p-value	Charateristics	No. (%)	p-value	Charateristics	No. (%)	p-value
χ² test			**Univariate analysis**			**Multivariate analysis**		
Cancer Risk			**Cancer Risk**			**Cancer Risk**		
Average	106(59.55)	0.0230	Average	106(59.55)	0.00217	Average	106(59.55)	0.0005
High	72(40.45)		High	72(40.45)		High	72(40.45)	
Age			**Age**			**Age**		
≤35	69(38.76)	0.2381	≤35	69(38.76)	0.2670	≤35	69(38.76)	0.6121
> 35	109(61.24)		> 35	109(61.24)		> 35	109(61.24)	
Racial identity			**Racial identity**			**Racial identity**		
White	148(83.15)	0.1074	White	148(83.15)	0.0985	White	148(83.15)	0.0568
Black	30(16.85)		Black	30(16.85)		Black	30(16.85)	
BMI			**BMI**			**BMI**		
Healthy	68(38.20)		Healthy	68(38.20)		Healthy	68(38.20)	
Overweight	72(40.45)	0.1067	Overweight	72(40.45)	0.2167	Overweight	72(40.45)	0.0340
Obese	38(21.35)		Obese	38(21.35)	0.0355	Obese	38(21.35)	0.0036
Menopausal Status			**Menopausal Status**			**Menopausal Status**		
Pre-	114(64.04)	0.7990	Pre-	114(64.04)	0.522	Pre-	114(64.04)	0.4491
Post-	64(35.96)		Post-	64(35.96)		Post-	64(35.96)	
Parity			**Parity**			**Parity**		
No	43(24.16)	0.4988	No	43(24.16)	0.0565	No	43(24.16)	0.1858
Yes	135(75.84)		Yes	135(75.84)		Yes	135(75.84)	

Figure 4-1. Preferential upregulation of TRIM37 in early-stage TNBC tumor and cancer-free breast tissue of Black women (BW). (A) The incidence rate of TNBC in female BW and White women (WW), excluding Hispanics, from 2010 to 2020 Surveillance, Epidemiology, and End Results (SEER) data. P value for years 2010–2020 ***$P < 2.2 \times 10^{-16}$, proportion test. (B, C) TNBC incidence (B) and mortality rate (C) per 100,000 for BW and WW TNBC patients stratified by age groups 15–39, 40–74, and above 75 years of age from 2010 to 2020. P values for the incidence and mortality rates in the different groups are ***$<2.2 \times 10^{-16}$, proportion test. (D, E) Box plot for racial identity-specific TRIM37 expression in Stage I (D; n = 71) and Stage (II–IV; n = 248) TNBC patients (GSE142731, Data ref: Saleh et al, 2021; PRJNA704957, Data ref: NCBI Sequence Read Archive PRJNA704957, 2021, and TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast). Stage I, *P = 0.029, and Stage II–IV, ns P = 0.058, unpaired t test.

118

The boxed areas span the first to the third quartile, with the central line representing the median expression changes for each group. Outliers from the boxplots are not displayed. The whiskers represent the 15th and 85th percentiles. (F) Kaplan–Meier survival curve showing overall survival for the BW and WW TNBC patients with high TRIM37 expression (GSE39004; Data ref: Tang et al, 2018", TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast), and GSE18229, Data ref: Prat et al, 2010). **P = 0.0011, log-rank test. The number of surviving patients at 0, 100, 200, and 300-month time points is indicated below the graph. (G) Box plot for TRIM37 expression in the normal breast tissue of BW (n = 119) and WW (n = 430) (GSE164641, Data ref: Marino et al, 2022; GSE111601, Data ref: Sun et al, 2018; GTEX, Data ref: The Genotype-Tissue Expression (GTEx), and TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast). **P = 0.0029, unpaired t test. The boxed areas span the first to the third quartile, with the central line representing the median expression changes for each group. Outliers from the boxplots are not displayed. The whiskers represent the 15th and 85th percentiles. (H) Association of TRIM37 expression in the cancer-free breast tissue with racial identity and age by chi-squared, univariate, and multivariate analyses. (I) Box plot of TRIM37 transcript level in the normal breast tissue of BW (n = 6) and WW (n = 103) ancestry confirmed by admixture analysis (TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast)). *P = 0.049, Wilcoxon test. The boxed areas span the first to the third quartile, with the central line representing the median expression changes for each group. Outliers from the boxplots are not displayed. The whiskers represent the 15th and 85th percentiles. (J) Box plot for TRIM37 transcript levels in breast tissues from women at average risk (n = 106) or high risk (n = 72) of developing breast cancer (GSE164641, Data ref: Marino et al, 2022). **P = 0.0022, unpaired t test. The boxed areas span the first to the third quartile, with the central line representing the median expression changes for each group. Outliers from the boxplots are not displayed. The whiskers represent the 15th and 85th percentiles. (K) Association of TRIM37 expression in breast tissues from women at average (n = 106) or high risk (n = 72) of developing breast cancer with clinicopathological variables (breast cancer risk, age, racial identity, BMI, menopausal status, and parity) by chi-squared, univariate, and multivariate analyses. Source data are available online for this figure.
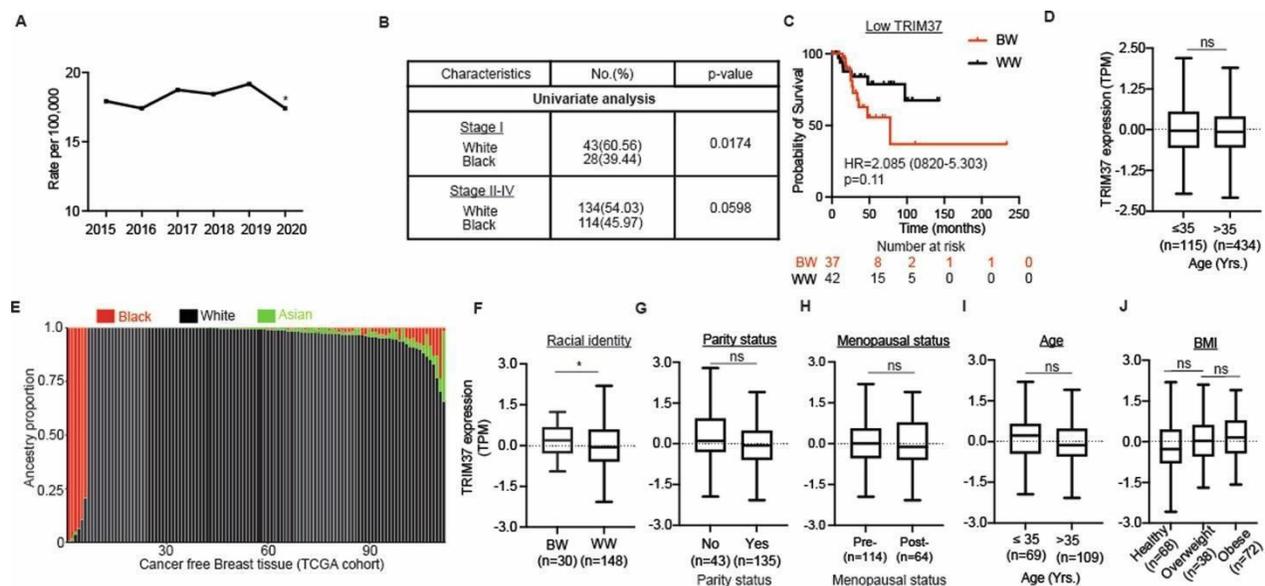
Supplementary Figure 4-1. Ancestry-specific differences in TNBC survival and incidence.
(A) Breast cancer incidence rate per 100,000 in women, excluding Hispanics, from 2015-2020
Surveillance, Epidemiology, and End Results (SEER) data. *P = 0.038, proportional test. The
number of samples is indicated. (B) Univariate analyses of relationships between TRIM37 and
racial identity in Stage I (n = 71) and Stage II–IV (n = 248) TNBC patients (GSE142731, Data ref:
Saleh et al, 2021; PRJNA704957, Data ref: NCBI Sequence Read Archive PRJNA704957, 2021,
and TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast)) (C) Kaplan–Meier
survival curve showing overall survival for the BW and WW TNBC patients with low TRIM37
expression (GSE39004, Data ref: Tang et al, 2018; TCGA, Data ref: The Cancer Genome Atlas
Program (TCGA-Breast), and GSE18229, Data ref: Prat et al, 2010). The number of surviving
patients at 0, 50, 100, 150, 200, and 250-month time points is indicated below the graph. ns
P = 0.11, log-rank test. The number of surviving patients at 0, 50, 100, 150, 200, and 250-month
time points are indicated. (D) Box plot for TRIM37 expression in the normal breast tissue by age,
below or equal to 35 (n = 115) vs. greater than 35 years (n = 434) (GSE164641, Data ref: Marino
et al, 2022"; GSE111601, Data ref: Sun et al, 2018; GTEX, Data ref: The Genotype-Tissue
Expression (GTEx), and TCGA, Data ref: The Cancer Genome Atlas Program (TCGA-Breast)).
ns P = 0.95, t test. The boxed areas span the first to the third quartile, with the central line
representing the median expression changes for each group. Outliers from the boxplots are not
displayed. The whiskers represent the 15th and 85th percentiles. (E) Estimated genetic ancestry
distribution for self-reported BW (n = 6) and WW (n = 103) in the TCGA cohort (Data ref: The
Cancer Genome Atlas Program (TCGA-Breast)). Each column represents an individual in the
cohort, and the estimated proportion of African, European, and Asian ancestry is shown on the y-
axis. (F–J) Box plot for TRIM37 expression in the breast tissue from women stratified by racial
identity (Black vs. White), parity (No vs. yes), menopausal status (Pre-vs. Post), age (below or
equal to 35 vs. greater than 35 years) and BMI (Healthy vs. overweight vs. obese) (GSE164641,
Data ref: Marino et al, 2022). For race, *P = 0.05, parity, ns P = 0.087, menopausal status, ns
P = 0.52, age, ns P = 0.3, and BMI, ns P = 0.17, t test. The number of samples is indicated. The
boxed areas span the first to the third quartile, with the central line representing the median
expression changes for each group. Outliers from the boxplots are not displayed. The whiskers
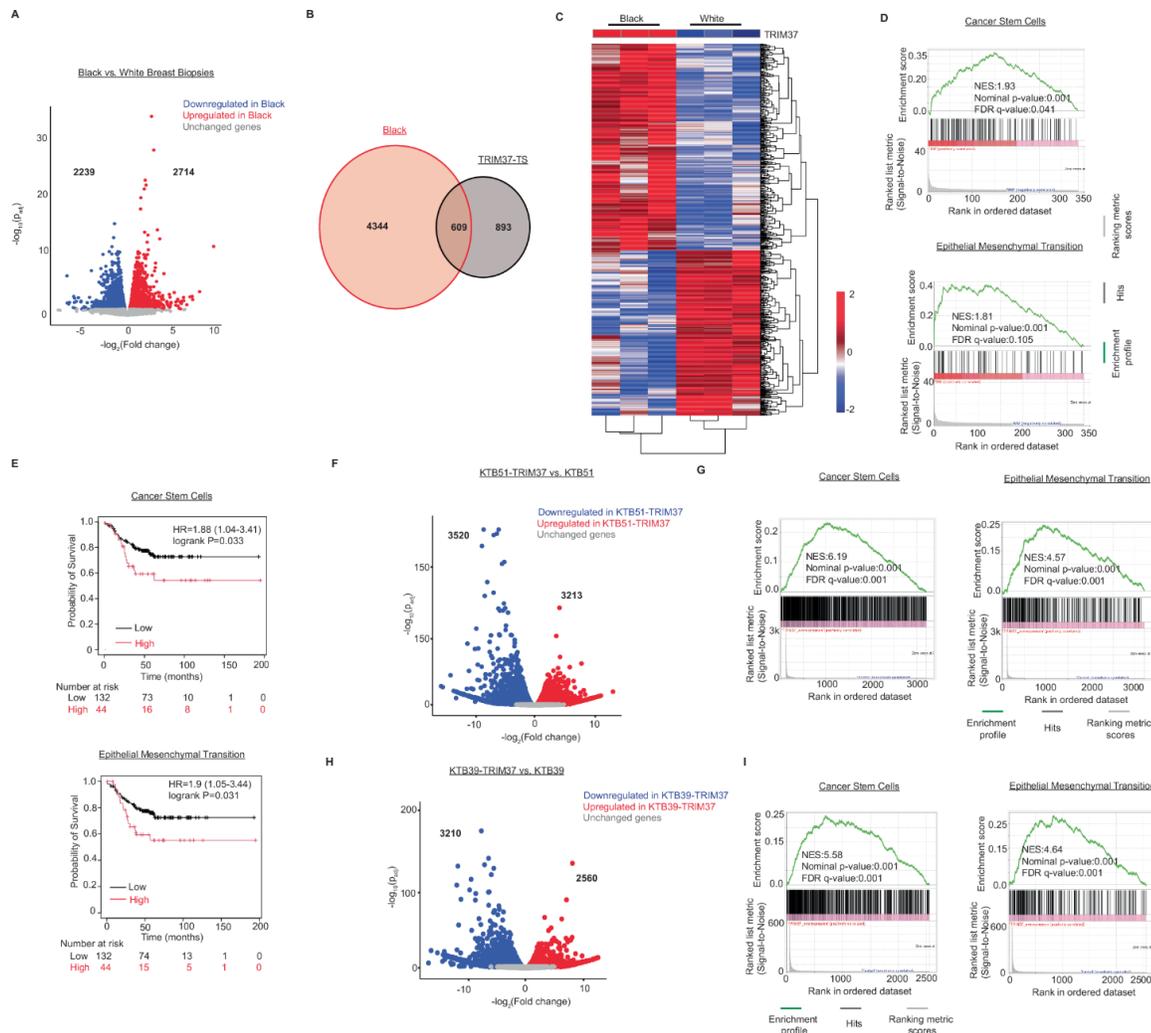represent the 15th and 85th percentiles.

African ancestry influences TRIM37-associated transcriptional signatures in

normal, cancer-free breast tissue

TRIM37 promotes tumorigenesis through its association with the polycomb

proteins, which are repressive complexes for chromatin-based gene regulation [19].

Thus, TRIM37 levels are expected to drive distinct gene expression changes

referred to as TRIM37-associated transcriptional signatures (TRIM37-TS). We

rationalized that the TRIM37-TS [19, 21] could better predict TNBC onset and

phenotype than TRIM37 alone. To test this idea, we investigated transcriptional profiles of the paraffin-archived normal breast tissue samples that were obtained from parous, premenopausal WW, and BW (The Susan G. Komen Tissue Bank; Indiana University Simon Comprehensive Cancer Center). These women were cancer-free, and donors were recruited under a protocol approved by the Indiana University Institutional Review Board (#101103097) and completed questionnaires with reproductive histories, demographic information, early life, and lifestyle habits.

The differential gene expression (DEG) analysis revealed 2714 upregulated and 2239 downregulated genes at an FDR cutoff of <0.1 in BW compared with WW breast tissues (Figs. 4-2A and Supplementary Fig.4-2A). Interestingly, among the 4953 DEGs, 609 were previously identified as TRIM37-TS (GSE136617, Data ref: Przanowski et al, 2020), revealing TRIM37-dependent molecular alterations in BW (Fig. 4-2B). Hierarchical clustering showed separation of transcriptomic profiles by ancestry into two distinct racial identity-specific clusters (Fig. 4-2C). To explore TRIM37-driven functional pathways in BW, we performed Gene Set Enrichment Analysis (GSEA) for upregulated genes, which identified the cancer stem cell (CSC) and epithelial to mesenchymal transitional (EMT) among the enriched gene sets (Supplementary Fig.4-2B and 4-2D). The integration of CSC and EMT gene signatures showed a strong association with poor survival for TNBC

patients with a hazard ratio of 1.88 (P = 0.033) and 1.9 (P = 0.031), respectively (Fig. 4-2E). To experimentally validate ancestry-specific TRIM37-TS, we stably over-expressed TRIM37 in immortalized breast epithelial cells derived from healthy women of African ancestry (KTB51 and KTB39; Supplementary Fig.4-2C). The ancestry of these cell lines was previously estimated using a panel of 41-ancestry-informative SNPs [23]. RNA-seq analysis revealed significant gene expression changes induced by TRIM37 overexpression in KTB51 (Figs. 4-2F and Supplementary Fig.4-2D) and KTB39 (Figs. 4-2H and Supplementary Fig.4-2F) cells relative to control cells. As expected, GSEA of the DEGs showed a positive enrichment of CSC and EMT pathways in TRIM37 overexpressing KTB51 (Figs. Supplementary Fig.4-2E and 4-2G) and KTB39 cells (Figs. Supplementary Fig.4-2G and 4-2I). Together, these results suggest that a higher level of TRIM37 primes normal breast tissue for neoplastic and premalignant changes.

Figure 4-2. African ancestry influences the *TRIM37*-associated transcriptional signatures in normal, cancer-free breast tissue. (A) Volcano plot illustrates differential gene expression in the cancer-free paraffin-archived breast tissue samples of parous and premenopausal and WW women (n = 3 biological replicates; ages <45). Differences in gene expression between BW and WW were tested using DESeq2 from Bioconductor. FDR < 0.1. Red are significantly upregulated genes (n = 2714), blue are significantly downregulated genes (n = 2239), and grey are genes not significantly changed (n = 11,448). (B) Venn diagram showing the overlap between differentially expressed genes in breast tissue from BW identified in (A) and TRIM37-regulated genes (GSE136617, Data ref: Przanowski et al, 2020). (C) Hierarchical clustering of median-centered TRIM37-TS in cancer-free BW and WW breast tissue (n = 3 biological replicates per group). Each colored line in the dendrogram identifies a different gene. (D) Enrichment plots for cancer stem cells (top) and EMT (bottom) gene signatures enriched in cancer-free BW (n = 3 biological replicates) breast tissue identified through GSEA analysis of RNA-seq data. Nominal P values were computed using the t test. (E) Kaplan–Meier analysis of survival in patients with low (black) or high (red) TRIM37-regulated CSC (top) and EMT (bottom) gene signatures enriched in cancer-free BW (n = 3 biological replicates) breast tissue identified in (B). For CSC signatures, *P = 0.033; for EMT signatures, *P = 0.031, log-rank test. The number of surviving patients at 0, 50, 100, and 150-month time points is indicated below the graph. (F) Volcano plot illustrates

differential gene expression in control and TRIM37 overexpressing KTB51 breast epithelial cells (n = 2 biological replicates per group). Differential gene expression between the groups was analyzed using DESeq2 from Bioconductor. FDR < 0.05. Red are significantly upregulated genes (n = 3213), blue are significantly downregulated genes (n = 3520), and grey are genes not significantly changed (n = 7113). (G) Enrichment plots for cancer stem cells (left) and EMT (right) gene signatures identified through GSEA analysis of RNA-seq data for TRIM37 overexpressing KTB51 cells. Nominal P values were computed using a t test. (H) Volcano plot illustrates differential gene expression in control (n = 2 biological replicates) and TRIM37 overexpressing (n = 2 biological replicates) KTB39 breast epithelial cells. Differential gene expression between the groups was analyzed using DESeq2 from Bioconductor. FDR < 0.05. Red are significantly upregulated genes (n = 2560), blue are significantly downregulated genes (n = 3210) and grey are genes not significantly changed (n = 7769). (I) Enrichment plots for cancer stem cells (left) and EMT (right) gene signatures identified through GSEA analysis of RNA-seq data for TRIM37 overexpressing KTB39 cells. Nominal P values were computed using a t test.
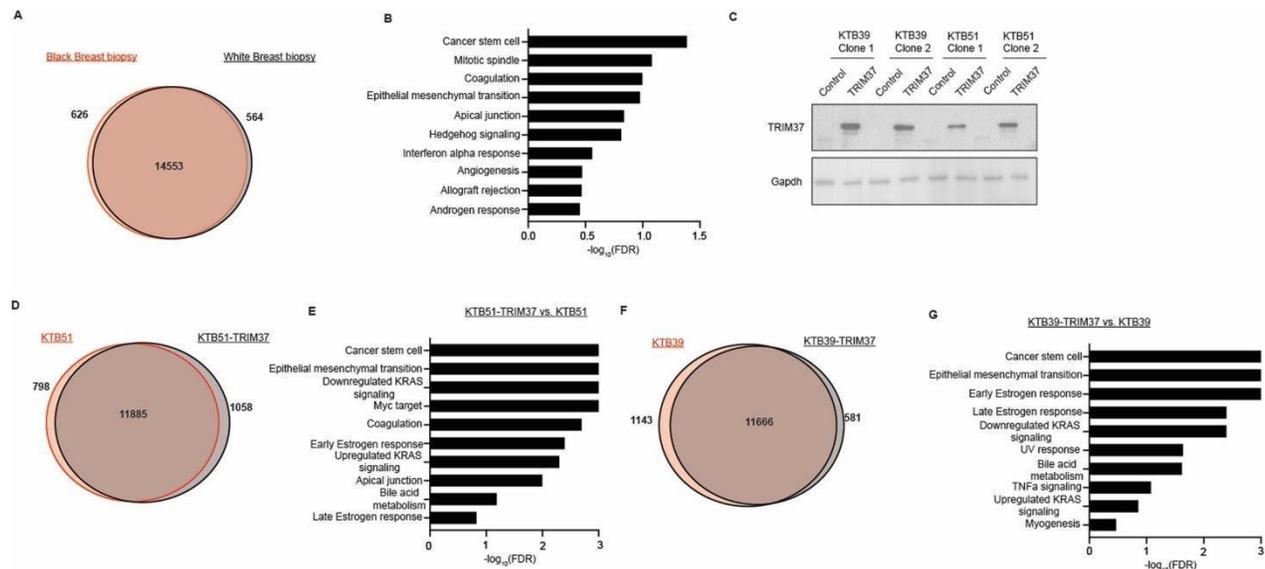
Supplementary **Figure 4-2. TRIM37 drives cancer stemness and epithelial to mesenchymal transition phenotype in Black breast epithelial cells.** (A) Venn diagram showing the overlap between DEGs in the cancer-free breast tissue samples of parous and premenopausal BW (n = 3; ages <45) and WW (n = 3; ages <45). (B) The top ten pathways enriched in the cancer-free breast tissue samples of BW relative to WW identified by GSEA are shown. (C) Immunoblot analysis of control and TRIM37 overexpressing KTB39 and KTB51 cells. Two clones for each cell line are shown. Gapdh was used as a loading control. (D) Venn diagram showing the overlap between differentially expressed genes in control and TRIM37 overexpressing KTB51 cells. (E) The top ten pathways significantly enriched in TRIM37 overexpressing KTB51 cells relative to control cells identified by GSEA are shown. (F) Venn diagram showing the overlap between differentially expressed genes in control and TRIM37 overexpressing KTB39 cells. (G) The top ten pathways enriched in TRIM37 overexpressing KTB39 cells relative to control cells identified by GSEA are shown.

A reporter-based screen identifies risk variant rs57141087 affecting TRIM37 promoter–enhancer interactions dominant in BW

We were surprised that TRIM37 was expressed at a higher level in the breast tissue of healthy BW than WW. To further investigate this finding, we annotated the single-nucleotide polymorphisms (SNPs) associated with TRIM37 for genomic coordinates (59,108,921–58,982,138 Mbp) available through the NHGRI (http://www.ncbi.nlm.nih.gov/snp/). For quality control, SNPs that deviated from Hardy–Weinberg equilibrium were excluded from further analyses. Of the 370 TRIM37-associated SNPs, we identified 154 variants predominant in BW with minor allele frequency (MAF) greater than or equal to 0.0003 (Fig. EV3A, Dataset EV1). Previous studies have shown that SNP location in promoters dictates gene expression and cancer susceptibility (Deng et al, 2017; Fagny et al, 2020). We, therefore, focused on nine SNPs located in the ~2 Kbp upstream of the transcription start site (TSS), including rs150880035 (C > T), rs57141087 (A > G), rs1029035382 (C > T), rs914034546 (G > A), rs1359804461 (A > C), rs904164828 (G > T), rs572032837 (T > G), rs568274755 (G > A), and rs1042165062 (G > C) (Fig. 4-3A). Figure 4-3B summarizes allelic frequencies in BW and WW for SNPs distributed in the TRIM37 gene upstream region. The linkage disequilibrium (LD) between SNPs located in Chromosome 17 (59,108,921–58,982,138 Mbp) indicated a block of strong LD (r2≥0.8) residing within ~2 Kbp of the 5′ promoter and intron 1 for (i)

rs8082544 with rs6503893, rs6503894, rs11868348, rs8067264, rs7503190, rs8081967, rs8182277, rs12948746, rs8081297, rs7502208, rs8072712, rs6416932rs114844088, rs144783389, and rs148648311; (ii) rs57141087 with rs142125398, rs6503895, rs8071291, rs75148295, rs73993836, rs57755522, rs8065740 and rs7224809, (iii) rs150880035 with rs115329428 and rs140241165 (Fig. Supplementary Fig.4-3B). To systematically identify functional SNPs among the variants distributed in the 5' gene upstream region of TRIM37, we carried out a small-scale candidate-based screen using a reporter-based promoter assay in HEK293T, MCF10a, and MDA MB 231 cells. Specifically, we cloned a 960-bp fragment (Chr17: 59,107,405-59,106,446) harboring either the reference or risk allele for each indicated SNP. As expected, we observed strong luciferase activity compared to the empty control (Supplementary Fig.4-3C), confirming the regulatory activity of the 5' gene upstream region of TRIM37. Among the SNP candidates, the risk allele for rs57141087 showed the maximum increase in promoter activity relative to the reference allele in the HEK293T, MCF10a, and MDA MB 231 cells (Fig. 4-3C), raising a possibility of genotype-dependent control of TRIM37 promoter activity.

We next interrogated the functional locus of rs57141087 to understand its association with TRIM37-driven breast cancer risk. The cis-expression quantitative trait loci (eQTL) analysis using 64 normal primary breast samples available through

GTEx showed the strongest association between rs57141087 and TRIM37 among all genes in the 126,784 bp window (Chr17:59,108,921–58,982,138; Fig. 4-3D). We found that the homozygous (G/G) or heterozygous (A/G) risk allele of rs57141087 strongly associated with TRIM37 expression relative to the homozygous (A/A) reference allele at rs57141087 (Fig. 4-3E). Furthermore, our analysis of previously published GWAS studies of breast cancer, including 2860 cases and 16,262 controls of African ancestry [34], identified a significant association between rs57141087 and TNBC risk at $P = 0.02$, with odd ratios of 1.096 (Supplementary Fig.4-3D). We next sought to understand the underlying mechanism by which rs57141087 modulates TRIM37 expression. rs57141087 is located 805 bp upstream of the TRIM37 TSS on chromosome 17 (Fig. 4-3F), which is highly enriched with enhancer marks, including histone H3 Lysine 4 mono-methylation (H3K4me1), histone H3 Lysine 27 acetylation (H3K27ac), deoxyribonuclease I (DNaseI) hypersensitive sites and histone H3 Lysine 4 trimethylation (H3K4me3) (Fig. 4-3G). We, therefore, hypothesized that the risk variant of rs57141087 could positively regulate TRIM37 expression by modulating enhancer activity. A luciferase-based reporter assay confirmed an enhancer activity of ~350 bp fragment containing rs57141087 upstream of the TRIM37 minimal promoter (Supplementary Fig.4-3E). Enhancer sequence with the risk allele at rs57141087 exhibited significantly stronger luciferase activity than that with the reference allele in

HEK293T cells (Fig. 4-3H). The inclusion of an additional 556 bp upstream

(Chr17: 59,107,406–59,107,961) did not accentuate the promoter activity relative to

353 bp fragment (Fig. 4-3H), indicating that the enhancer activity is limited to the

59,107,053–59,107,405 regions in the TRIM37 5' regulatory region.

Previous studies have identified looping between distant chromatin segments as an

important feature of enhancer function during active transcription [36-40]. We,

therefore, measured the TRIM37 enhancer–promoter contact in the context of

rs57141087 through a cell-free looping assay and quantitated by looping index [25].

As shown in Fig. 4-3I, we observed a significantly higher looping index for the

construct harboring risk allele relative to the reference allele at rs57141087.

Substituting a nuclear extract with BSA inhibited looping, indicating that promoter–

enhancer looping depends on the protein factors (Fig. 4-3I). A prior treatment with

DNase abolished the interactions (Fig. 4-3I). Together, these results indicate that

the risk allele of the rs57141087 favors stronger TRIM37 promoter activity by

facilitating promoter–enhancer interactions.
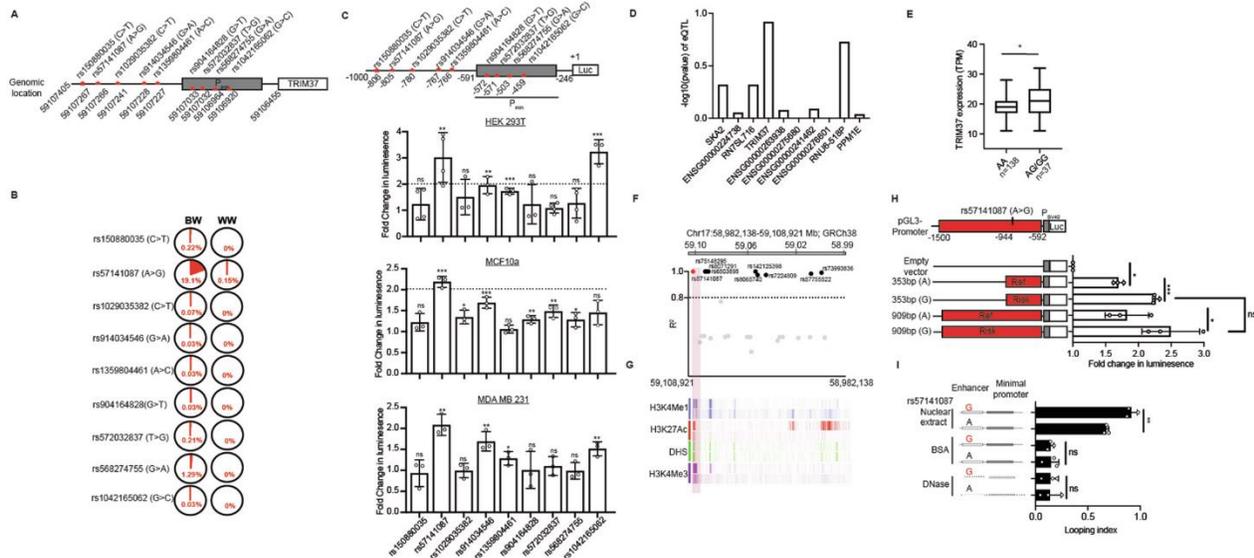
Figure 4-3. rs57141087 modulates enhancer activity to regulate *TRIM37* expression. (A, B) A schematic view of the (Chr17: 59,107,405-59,106,446) region (A) and pie charts (B) showing the distribution of risk (red) and reference (black) alleles in BW and WW from the NCBI SNP database. The percentage of each haplotype in Black and White populations is indicated. (C) Luciferase reporter assays measure the promoter performance of the promoter harboring risk alleles for different SNPs (top) in HEK293T, MCF10a, and MDA MB 231 cells. The data are normalized to the reference allele and presented as a fold change in luminescence. For HEK293T; rs150880035 ns P = 0.463, rs57141087, **P = 0.005, rs1029035382, ns P = 0.268, rs914034546, **P = 0.007, rs1359804461, ***P = 0.0003, rs904164828, ns P = 0.621, rs572032837, ns P = 0.398, rs568274755, ns P = 0.383, rs1042165062, ***p = 0.001; For MCF10a rs150880035, ns P = 0.133, rs57141087, ***P = 9.4 × 10−5, rs1029035382, *P = 0.021, rs914034546, ***P = 7.72 × 10−3, rs1359804461, ns P = 0.330, rs904164828, **P = 0.005, rs572032837, **P = 0.005, rs568274755, *P = 0.045, rs1042165062, ns P = 0.053; For MDA MB 231; rs150880035, ns P = 0.731, rs57141087, **P = 0.002, rs1029035382, ns P = 0.974, rs914034546, **P = 0.007, rs1359804461, *P = 0.036, rs904164828, ns P = 0.974, rs572032837, ns P = 0.469, rs568274755, ns P = 0.928, rs1042165062, **P = 0.006, unpaired t test. Data are mean ± SD of biological replicates, n ≥3/group. (D) rs57141087 eQTL analysis of genes within 126,784-bp (Chr17:59,108,921–58,982,138) window using data for breast tissue from GTEx. (E) eQTL analysis demonstrating the correlation between rs57141087 genotype and TRIM37 expression in samples from 1000 genome RNA-seq dataset (n = 175, Yoruba and UTAH population, Data ref: Lappalainen et al, 2013). *P = 0.041, t test. The boxed areas span the first to the third quartile, with the central line representing the median expression changes for each group. Outliers from the boxplots are not displayed. The whiskers represent the 15th and 85th percentiles. (F) The LocusZoom plot of LD for rs57141087 and its LD-SNPs in African (Yoruba, Luhya, Gambia, Mende, Esan, Americans of African Ancestry, and African Caribbean) population from the 1000 Genome project data. (G) The epigenetic heatmap for H3K4me1, H3K27Ac, DHSs, and H3K4me3 for the (Chr17:59,108,921–58,982,138) region harboring rs57141087. The data was derived from the Cistrome Data Browser. (H) Luciferase reporter assays measuring the enhancer activity of DNA fragment (Chr17: 59,107,961–59,107,053) harboring risk alleles for rs57141087 in HEK293T cells. The data are normalized to the vector only set to 1 and presented as a fold

change in luminescence. For 353 bp, ***P = 4.8*10−4, 909 bp, *P = 0.015, 353 bp vs. 909 bp, ns P = 0.418, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (I) Looping assay measures the promoter–enhancer interactions for in vitro synthesized DNA fragments (Chr17: 59,107,405-59,106,446) with risk or reference alleles. For Nuclear extract, **P = 0.003, BSA, $^{ns\ P}$ = 0.947, ns DNase, P = 0.965, unpaired t test. Data are mean±SD of biological replicates, n = 3/group.
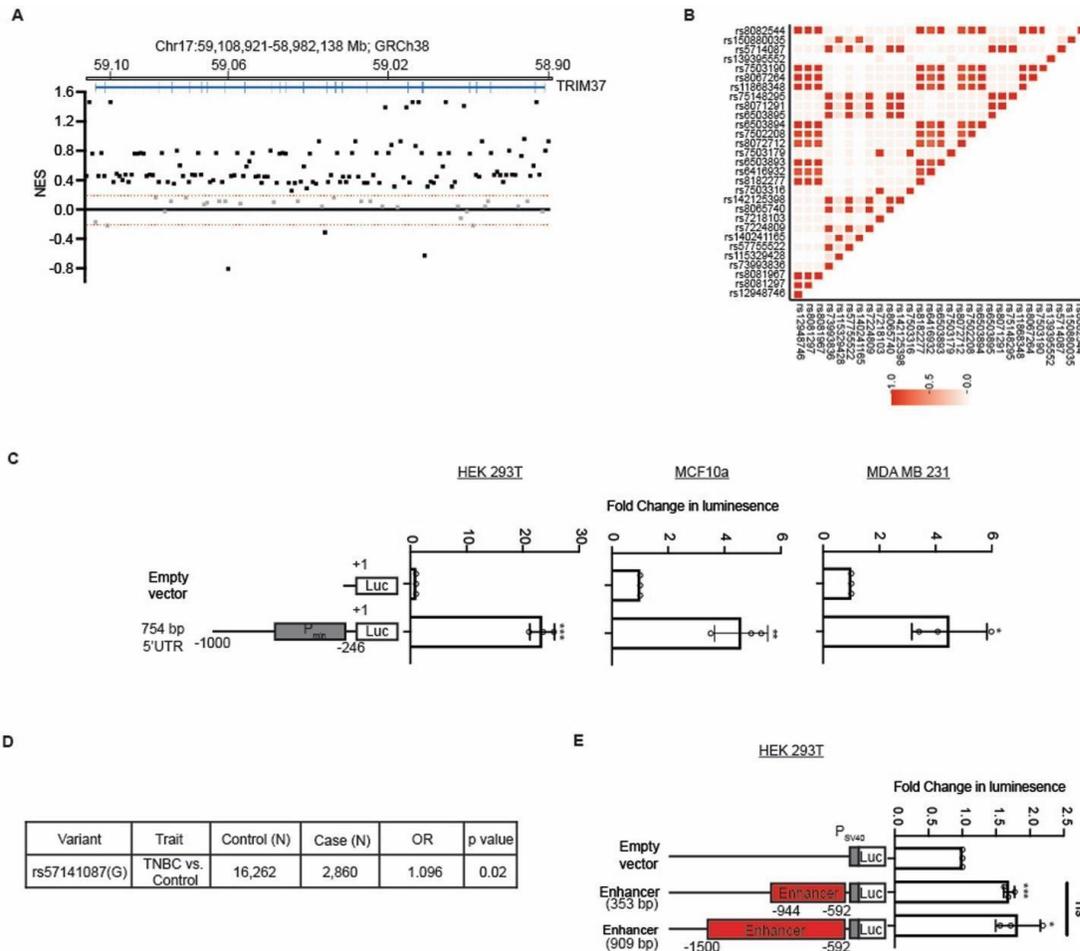
Supplementary Figure 4-3. rs5714087 increases *TRIM37* expression through modulation of enhancer function. (A) Distribution of SNPs across the TRIM37 gene (Chr17:59,108,921–58,982,138 Mb) and their association with TRIM37 expression in the breast tissue from GTEx. (B) Linkage disequilibrium (r2) heatmap of SNPs in TRIM37 gene in African (Yoruba, Luhya, Gambia, Mende, Esan, Americans of African Ancestry, African Caribbean) population from the 1000 Genome project data. (C) Luciferase reporter assays measure the promoter performance of the TRIM37 minimal promoter (Chr17: 59,107,052–59,106,707) in HEK293T, MCF10a, and MDA MB 231 cells. For HEK293T, ***P = 6.05*10-5, MCF10a, **P = 0.008, and MDA MB 231, *P = 0.012, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (D) TNBC risk association with the risk allele of rs57141087 using GWAS data for African ancestry (Jia et al, 2024). (E) Luciferase reporter assays measuring the enhancer activity of DNA fragments (Chr17: 59,107,961–59,107,053) in HEK293T cells. The data is normalized to the vector only set to 1 and presented as a fold change in luminescence. of the different fragments in HEK293T. For 353 bp,

***P = 0.0001, 909 bp, *P = 0.012, and 353 bp vs. 909 bp, P = 0.539, unpaired t test. Data are mean±SD of biological replicates, n = 3/group.

The risk allele rs57141087 upregulates TRIM37 via NRF1

To determine the effect of rs57141087 at its endogenous locus, we leveraged CRISPR/Cas9 to generate a clonal KTB51 cell line harboring risk allele G instead of A allele of rs57141087, hereafter referred to as 51-G. We obtained three heterozygous clones for the risk allele of rs57141087 that were confirmed by risk SNP-specific restriction digestion (Supplementary Fig.4-4A) and next-generation sequencing of the target region (Fig. 4-4A). To confirm the functional integrity of the locus, the PolII recruitment to TRIM37 promoter was evaluated in 51-G and parental KTB51 cells (referred to as 51-A). The chromatin immunoprecipitation (ChIP) analysis revealed significantly higher PolII binding at the TRIM37 promoter in 51-G than in 51-A (Fig. 4-4B). Given that SNPs in enhancer regions can function by modulating transcription factor (TF) binding, we evaluated the protein complexes binding capabilities of TRIM37 enhancer DNA fragment in the context of rs57141087 risk and reference allele by electrophoretic mobility shift assay (EMSA). The result showed that the G allele has ~3.92-fold stronger binding capabilities than the A allele (Figs. 4-4C and Supplementary Fig.4-4B). As expected, prior treatment of the probe with the DNase abolished the signal (Figs.4-4C and Supplementary Fig.4-4B). Prompted by these findings, we next scanned the rs57141087 locus for potential TF binding sites, which showed that the risk allele,

131

but not the reference allele, of rs57141087, creates the unique binding sites for

Growth factor independent-1 (GFI1), Paired Box 5 (PAX5) and Nuclear

Respiratory Factor 1 (NRF1) in the enhancer region (Figs.4-4D and Supplementary

Fig.4-4C,D). GFI1 is a transcriptional repressor critical in myeloid and lymphoid

differentiation and lymphocyte effector functions [41]. PAX5 regulates early

development and induces transcriptional signatures associated with neoplastic

transformations [42]. NRF1 homodimerizes to regulate the expression of metabolic

genes and nuclear genes required for DNA transcription and replication [43, 44].

Interestingly, the shRNA-mediated knockdown of PAX5 and GFI1 in 51-G and 51-

A did not alter TRIM37 levels (Supplementary Fig.4-3E,F), excluding their role in

the transcriptional regulation of TRIM37 via the SNP rs57141087. In contrast,

NRF1 knockdown in 51-G significantly decreased TRIM37 levels (Fig. 4-4E). No

significant change in TRIM37 levels was observed in 51-A following NRF1

knockdown (Fig. 4-4E).

We next interrogated the physical interaction of NRF1 with the enhancer

region harboring the risk allele of SNP rs57141087. The ChIP assay showed a

significantly higher enrichment of NRF1 in the enhancer region for 51-G cells

relative to 51-A cells (Fig. 4-4F). In concert, the protein binding affinities of DNA

fragment harboring risk allele of rs57141087 showed a dramatic decrease when

incubated with the nuclear extract from NRF1 knockdown cells relative to control cells (Figs. 4-4G and Supplementary Fig.4-4G).

We argued that disrupting the NRF1 binding in the TRIM37 locus would interfere with promoter–enhancer interactions. To test this idea, we carried out an *in vitro* looping assay using a 960 bp fragment harboring either a risk or reference allele in the enhancer region and minimal promoter with wild-type or mutated NRF1 binding sites. As shown in Supplementary Fig.4-4H, the loss of NRF1 binding in the TRIM37 locus significantly lowered the looping index for the construct harboring reference allele relative to the risk allele at rs57141087. As expected, substituting a nuclear extract with BSA or prior treatment with DNase abolished the interactions (Supplementary Fig.4-4H). Encouraged by these results, we performed quantitative chromosome conformation capture (3C-qPCR) assays to examine rs57141087 interaction with the TRIM37 minimal promoter *in vivo* [26]. Using an anchor primer in the minimal promoter and probes in the region harboring rs57141087, we observed a strong chromatin interaction between the rs57141087 and the TRIM37 promoter in 51-G cells but not in 51-A cells (Fig.4-4H). Significantly, the NRF1 knockdown is accompanied by a decreased PolII enrichment at the TRIM37 promoter relative to the control 51-G cells (Fig.4-4I). Together, these results confirm a loop structure formation between the SNP rs57141087-containing

enhancer region and the TRIM37 promoter, which regulates TRIM37 promoter activity.

To further explore the transcriptional effects of risk variant rs57141087, we carried out RNA-seq analysis in 51-A and 51-G. Two biological replicates were included in each group, and high correlations were observed between the replicates (Supplementary Fig.4-4I,J; r = 0.96 for 51-A and r = 0.95 for 51-G). As expected, TRIM37 was expressed at a significantly higher level in 51-G than in 51-A, as measured by qRT-PCR (Supplementary Fig.4-4K). We identified 807 and 837 genes that were significantly upregulated and downregulated in 51-G, respectively (Fig. 4-4J). Gene Set Enrichment Analysis (GSEA) analysis identified CSC and EMT terms highly enriched in 51-G upregulated genes (Supplementary Fig.4-4L and 4-4K) in concordance with TRIM37 function in tumorigenesis and breast cancer progression. Together, our results indicate that risk variant rs57141087 brings enhancer in contact with TRIM37 promoter in cis through NRF1 binding to induce precise gene expression changes triggering neoplastic transformation and breast cancer progression (Fig. 4-4L).
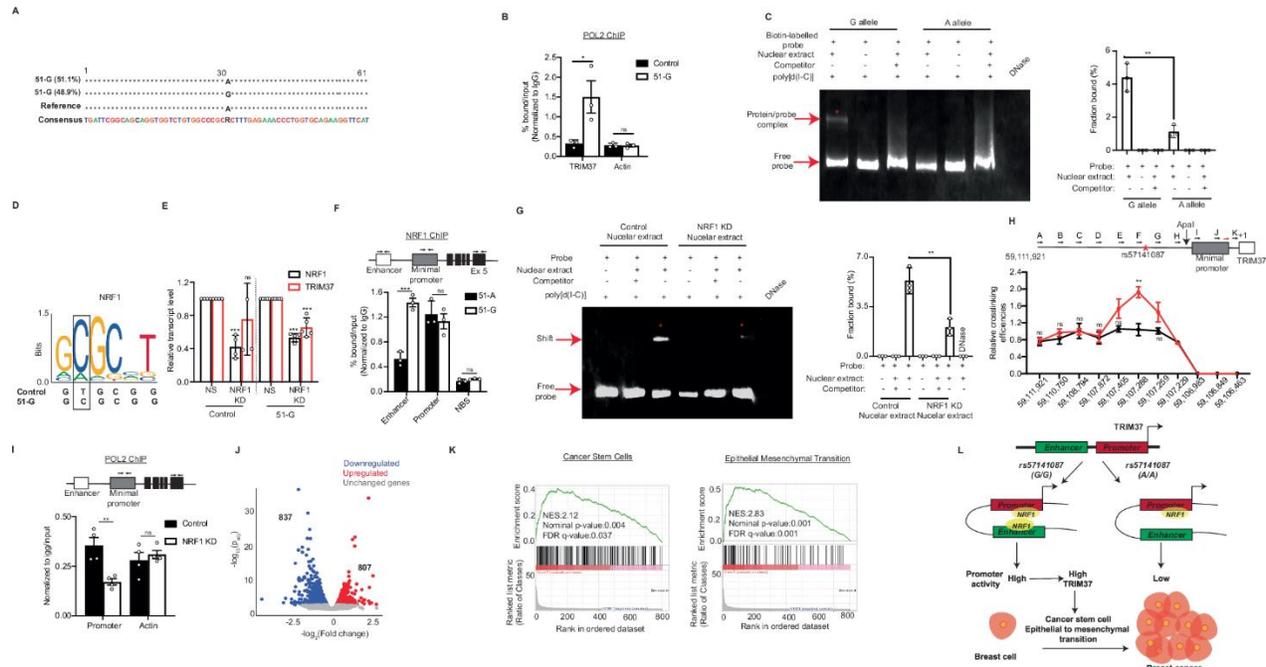
Figure 4-4. rs57141087 modulates TRIM37 promoter-enhancer interactions through NRF1 binding. (A) A schematic summary for amplicon sequencing of A- and G allele for rs57141087 inserted through CRISPR/Cas9 technology in KTB51 breast epithelial cells. (B) ChIP assay monitoring PolII recruitment on TRIM37 promoter in KTB51 cells with A or G allele for rs57141087. Actin is used as an endogenous control. For TRIM37, *P = 0.045, and Actin, ns P = 0.889, unpaired t test. Data are mean ± SD of biological replicates, n = 3/group. (C) Immunoblot for biotin-labeled TRIM37 enhancer fragment with risk (G allele) and reference (A allele) incubated with HEK293T nuclear extract and competitor (unlabeled probe) as indicated. The bound and free DNA fragments are indicated with red arrows. The concentrations of nuclear extract (5 µg), Biotin-labeled probes (0.2 pmol), competitor (10 pmol), and poly [d(I-C)] (50 ng/µl) were indicated at the top. The representative image from one experiment is shown (left), and the shift in the band is indicated (red asterisk, *) and quantitated (right). **P = 0.004, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (D) The risk G allele of rs57141087 harboring NRF1 binding motif in the enhancer region of TRIM37 (Chr17: 59,107,267–59,107,262) is shown. (E) qRT-PCR monitoring NRF1 and TRIM37 levels in KTB51 with A or G allele for rs57141087. Gapdh is used as an endogenous control. For reference allele NRF1, ***P = 0.0001, and TRIM37, ns P = 0.230; For risk allele NRF1, ***P = 2.68 × 10−8, and TRIM37, ***P = 0.0001, unpaired t test. Data are mean ± SD of biological replicates, n ≥ 4/group. (F) ChIP assay monitoring NRF1 recruitment on TRIM37 locus in 51-G and 51-A cells. NRF1 non-binding site (NBS) is used as a negative control. The primers used for the qPCR analysis in the TRIM37 locus are indicated on Top. For Enhancer, ***P = 0.0007, Promoter, ns P = 0.544, and NBS, ns P = 0.149, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (G) Immunoblot for biotin-labeled TRIM37 enhancer fragment incubated with nuclear extract from control or NRF1 knockdown HEK293T cells and unlabeled probe (competitor) as indicated. The bound and free DNA fragments are indicated with red arrows. The concentrations of nuclear extract (5 µg), Biotin-labeled probes (0.2 pmol), competitor (10 pmol), and poly [d(I-C)] (50 ng/µl) were indicated at the top. The representative image from one experiment is shown

(left), and the shift in the band is indicated (red asterisk, *) and quantitated (right). **P = 0.006, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (H) 3C-qPCR analysis across the ~5500-bp TRIM37 locus harboring risk and reference allele of SNP rs57141087 in 51-G and 51-A cells, respectively. The schematic shows the TRIM37 gene structure with PCR primers used for ChIP-qPCR assay (top). For (A), ns P = 0.780, (B), ns P = 0.206, (C), ns P = 0.918, (D), ns P = 0.814, (E), ns P = 0.069; (F), **P = 0.002, (G), ns P = 0.104, (H), ns P = 0.247, and (I–K), not applicable, paired t test. Data are mean ± SD of biological replicates, n = 3/group. (I) ChIP assay monitoring PolII recruitment on TRIM37 promoter in control and NRF1 knockdown KTB51 cells. Actin is used as an endogenous control. The primers used for the qPCR analysis in the TRIM37 locus are indicated on top. For Promoter, **n = 0.005, and Actin, nsn = 0.551, unpaired t test. Data are mean ± SD of biological replicates, n = 4/group. (J) Volcano plot illustrates differential gene expression in KTB51 with A or G allele for rs57141087 (n = 2 biological replicates). Differential gene expression between the groups was analyzed by DESeq2 Bioconductor. FDR < 0.05. Red are significantly upregulated genes (n = 807), blue are significantly downregulated genes (n = 837), and grey are genes not significantly changed (n = 11,628). (K) Enrichment plots for cancer stem cells (left) and EMT (right) gene signatures identified through GSEA analysis of RNA-seq data. Nominal P values were computed using a t test. (L) Graphical representation of the TRIM37 promoter regulation by the risk allele of rs57141087 and its effect on breast cancer progression.
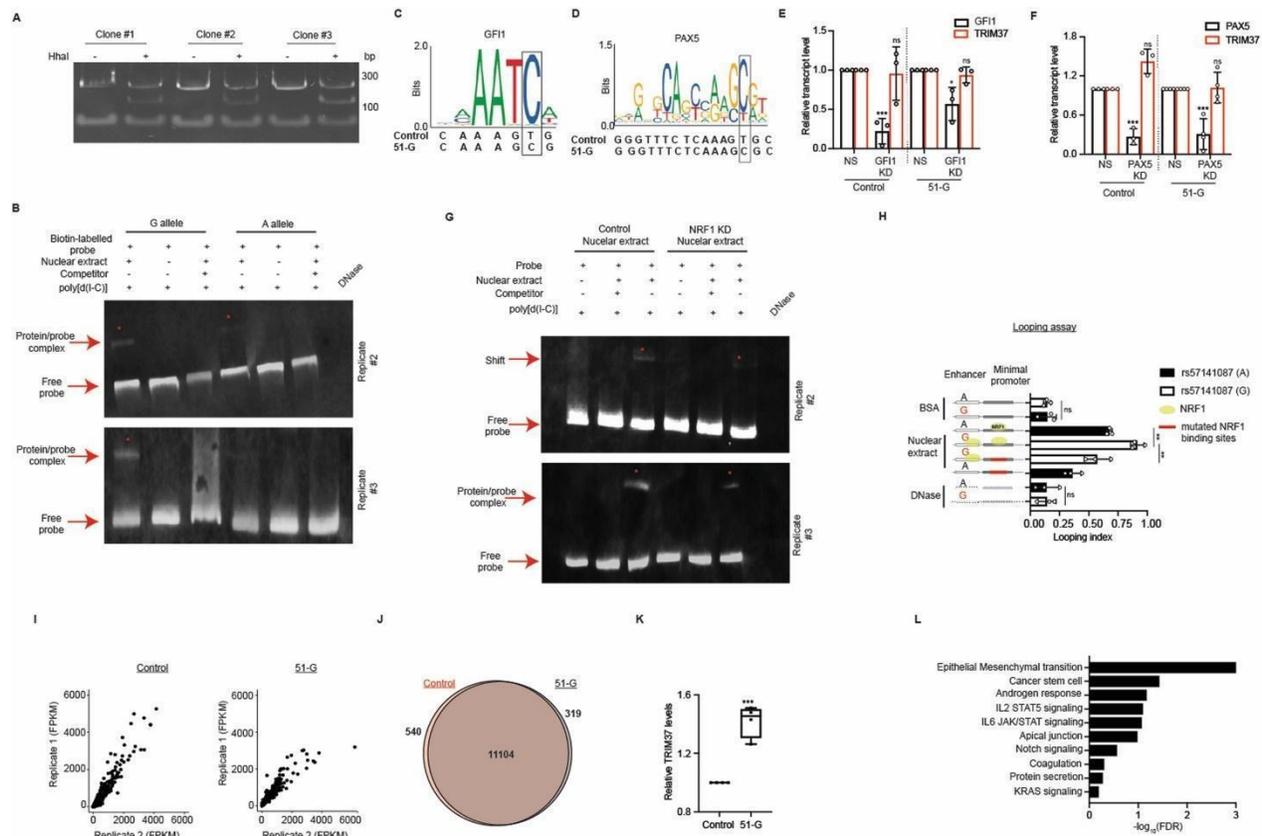
Supplementary Figure 4-4. NRF1 binds uniquely to the risk variant of rs5714087. (A) Representative polyacrylamide gel indicates a distinct HhaI digestion pattern for KTB51 clones harboring rs57141087 reference (A) and risk (G) alleles. The data for three different clones is shown. (B) Immunoblot for biotin-labeled TRIM37 enhancer fragment with risk (G allele) and

reference (A allele) incubated with HEK293T nuclear extract and competitor (unlabeled probe) as indicated. The bound and free DNA fragments are indicated with red arrows, and the shift in the band is indicated (red asterisk, *). The concentrations of nuclear extract (5 µg), Biotin-labeled probes (0.2 pmol), competitor (10 pmol), and poly [d(I-C)] (50 ng/µl) were indicated at the Top. The representative images from two different experiments (Replicate #1–2) are shown. (C, D) The risk G allele of rs57141087 harboring GFI1 (C) and PAX5 (D) binding motif in the enhancer region of TRIM37 (Chr17: 59,107,267–59,107,262) is shown. (E, F) qRT-PCR monitoring TRIM37 levels in GFI1 (E) and PAX5 (F) knockdown in KTB51 cells with A or G allele for rs57141087. Gapdh is used as an endogenous control. For GFI1 knockdown, GFI1 (A), ***P = 0.001, TRIM37 (A), ns P = 0.831, GFI1 (G), *P = 0.024, and TRIM37 (G), ns P = 0.345. For PAX5 knockdown, PAX5 (A), ***P = 0.0005, TRIM37 (A), ns P = 0.187, PAX5 (G), ***P = 0.001, and TRIM37 (G), ns P = 0.874, unpaired t test. Data are mean±SD of biological replicates, n ≥3/group. (G) Immunoblot for biotin-labeled TRIM37 enhancer fragment incubated with nuclear extract from control or NRF1 knockdown HEK293T cells and unlabeled probe (competitor) as indicated. The bound and free DNA fragments are indicated with red arrows, and the shift in the band is indicated (red asterisk, *). The concentrations of nuclear extract (5 µg), Biotin-labeled probes (0.2 pmol), competitor (10 pmol), and poly [d(I-C)] (50 ng/µl) were indicated at the Top. The representative images from two different experiments (Replicate #1–2) are shown. (H) A looping assay measuring the promoter–enhancer interactions for in vitro synthesized DNA fragments (Chr17: 59,107,405-59,106,446) with risk or reference alleles or mutated NRF1 binding motifs in the promoter was used. For BSA, ns P = 0.947, NRF1, **P = 0.003, NRF1 vs mutated NRF1, **P = 0.009, and DNase, ns P = 0.965, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (I) Correlation plots of the FPKM read from RNA-seq analysis for each replicate are shown for control (Left) and 51-G (Right) cells. (J) Venn diagram showing the overlap between differentially expressed genes in control and 51-G cells. (K) qRT-PCR analysis monitoring TRIM37 in control and 51-G cells. Gapdh is used as an endogenous control. ***P = 0.0003, unpaired t test. Data are mean ± SD of biological replicates, n = 4/group. The boxed areas span the first to the third quartile. The whiskers represent the 15th and 85th percentiles. (L) The top ten pathways enriched in 51-G relative to control cells identified by GSEA are shown.

TRIM37 drives oncogenic transformation in ancestry-mapped immortalized, healthy breast epithelial cells

Next, we assessed the tumorigenic capabilities of breast epithelial cells from BW in the context of TRIM37. To this end, we stably over-expressed TRIM37 in KTB51 and a RAS expressing premalignant derivative of KTB51, referred to as K51t (Fig. 4-5A). Among the different KTB51 derivatives tested, the ectopic expression of TRIM37 potentiated the proliferation and colony formation in KTB51

and K51t cells (Fig. 4-5B,C). Inspired by the results of the in-silico analysis (Fig. 4-2), we performed a series of phenotypic analyses to determine the tumorigenic and metastatic potential of TRIM37-transformed KTB51 cells. The wound-healing assay showed rapid wound closure within four hrs. for TRIM37-transformed KTB51 (K51-TRIM37) and KTB51t (K51t-TRIM37) cells relative to control cells (Figs. 4-5D and Supplementary Fig.4-5A). A higher number of mammospheres were formed by KTB51 derivatives expressing TRIM37, with K51t-TRIM37 cells showing the greatest increase in self-renewal phenotype (Figs. 4-5E and Supplementary Fig.4-5B). We next subjected mammospheres to flow cytometry for CD44/CD24 and EpCAM expression to evaluate the CSC and EMT phenotype, respectively. As expected, ~4.42- and ~1.68-fold increase in CD44high/CD24low and EpCAM were observed for K51t-TRIM37 than in KTB51 cells, respectively (Figs. 4-5F,G and Supplementary Fig.4-5C,D). In concert with FACS results, the qRT-PCR analysis of the mammospheres confirmed significantly increased expression of EMT gene signatures in KTB51 derivatives (Fig. 4-5H). In parallel, we generated TRIM37-expressing derivatives of white breast epithelial KTB37 cells (Fig. EV5E), which also showed increased cellular proliferation (Supplementary Fig.4-5F,G). Consistently, we find that TRIM37 overexpression is also sufficient to potentiate the CSC and EMT phenotype in KTB37 cells (Supplementary Fig.4-5H–M), ruling out additional ancestry-specific genetic or

epigenetic factors that could contribute to the tumorigenic potential of TRIM37. We next directly investigated the oncogenic potential of TRIM37-transformed KTB51 derivatives through growth in soft agar. Notably, the ectopic expression of TRIM37 was sufficient to transform KTB51 cells, which showed a 2.42-fold increase in colonies for KTB51-TRIM37 cells relative to KTB51 cells (Fig. 4-5I). The K51t-TRIM37 showed a maximum number of colonies in soft agar (Fig. 4-5I). Finally, the ectopic expression of TRIM37 enabled xenograft tumor formation in mice, with the K51t-TRIM37 tumors showing accelerated growth relative to control cells (Fig. 4-5J). The tumor growth was confirmed by hematoxylin and eosin (H&E) staining (Fig. 4-5K), the gross tumor tissue isolated postmortem (Fig. Supplementary Fig.5N), and a high proliferative index as determined by Ki67 staining (Fig. 4-5L). The immunohistochemistry (IHC)-based staining for the Estrogen-α receptor (ER), progesterone-receptor (PR), and human epidermal growth factor receptor 2 (HER2) showed no significant staining in K51t-TRIM37 tumors relative to the positive control (Fig. 4-5M–O). Collectively, these results confirmed that high TRIM37 levels in the early stages of transformation can inherently wire breast epithelial cells to accelerated TNBC growth and tumorigenesis.
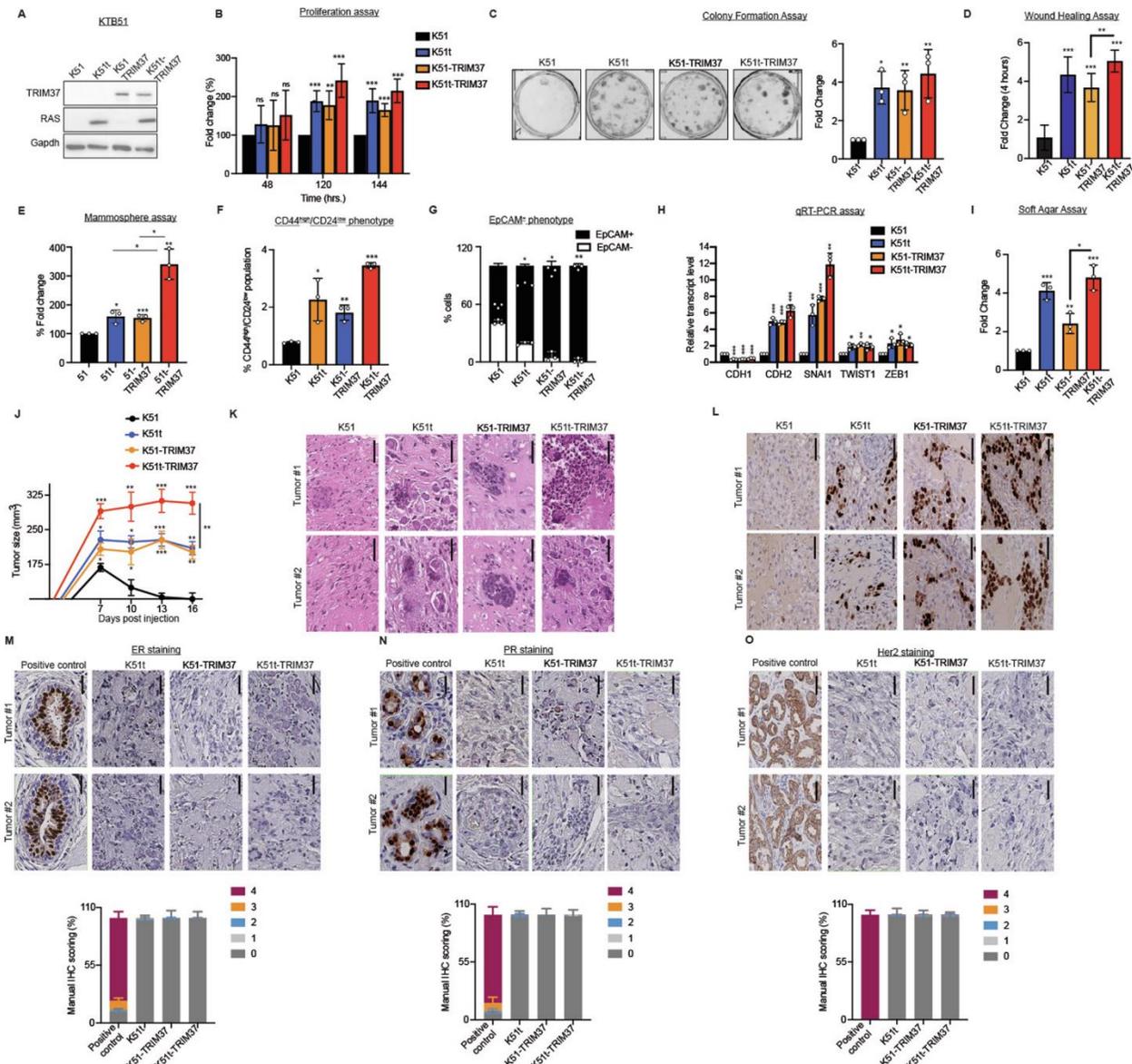
Figure 4-5. TRIM37 drives cancer stem cell and EMT phenotype to enhance breast cancer initiation and growth. (A) Immunoblots in KTB51 and TRIM37 derivatives of RAS-transformed KTB51 (K51t) cells. Gapdh was the loading control. (B) Relative cell growth for KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells at indicated times. For 48 h, K51t, ns P = 0.234, K51-TRIM37, ns P = 0.413, and K51t-TRIM37, ns P = 0.108. For 120 h., K51t, ***P = 8.96×10^{-5}, K51-TRIM37, **P = 0.002, and K51t-TRIM37, ***P = 8.64×10^{-5}. For 144 h., K51t, ***P = 0.0002, K51-TRIM37, ***P = 2.61×10^{-5}, and K51t-TRIM37, ***P = 3.10×10^{-5}, unpaired t test. Data are mean ± SD of biological replicates, n = 5/group. (C) Representative bright-field images after crystal violet staining showing the growth of KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. The colonies were quantified (right). For K51t, *P = 0.012, K51-TRIM37, **P = 0.005, and K51t-TRIM37, P = **0.009, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (D) The relative migratory abilities for KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells were quantitated after 4 h. For K51t, ***P = 3.3×10^{-5}, K51-TRIM37, ***P = 6.44×10^{-5}, and K51t-TRIM37, ***P = 5.26×10^{-7}; K51t vs. K51t-TRIM37,

ns P = 0.138, K51-TRIM37 vs. K51t-TRIM37, **P = 0.005, unpaired t test. Data are mean ± SD of biological replicates, n = 6/group. (E) Quantitation of solid mammospheres formed by KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. For K51t, *P = 0.012, K51-TRIM37, ***P = 0.001, and K51t-TRIM37, **P = 0.005; K51t vs. K51t-TRIM37, *P = 0.024, K51-TRIM37 vs. K51t-TRIM37, *P = 0.018, unpaired t test. Data are mean±SD of biological replicates, n = 6/group. (F, G) FACS analysis of CD24 and CD44 (F) and EpCAM (G) in KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells derived from mammospheres in (E). For CD44high/CD24low in K51t, *P = 0.026, K51-TRIM37, **P = 0.003, and K51t-TRIM37, ***P = 1.31×10^{-6}. For EpCAM+ in K51t, *P = 0.011, K51-TRIM37, *P = 0.011, and K51t-TRIM37, **P = 0.003, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (H) qRT-PCR analysis of EMT markers in KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells derived from mammospheres in (E). Gapdh is used as an endogenous control. For CDH1 in K51t, ***P = 8.08×10^{-5}, K51-TRIM37, ***P = 1.70×10^{-5}, and K51t-TRIM37, ***P = 0.0003; CDH2; K51t, ***P = 9.11×10^{-5}, K51-TRIM37, ***P = 0.0004, and K51t-TRIM37, ***P = 0.0002, For SNAI1 in K51t, **P = 0.003, K51-TRIM37, ***P = 2.8×10^{-6}, and K51t-TRIM37, **P = 0.005. For TWIST1 in K51t, *P = 0.012, K51-TRIM37, **P = 0.0014, and K51t-TRIM37, *P = 0.022. For ZEB1 in K51t, *P = 0.019, K51-TRIM37, *P = 0.013, and K51t-TRIM37, *P = 0.020, unpaired t test. Data are mean ± SD of biological replicates, n = 3/group. (I) Quantitation of colonies formed in soft agar by KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. For K51t, ***P = 0.0002, K51-TRIM37, **P = 0.009, and K51t-TRIM37, ***P = 0.0005; K51t vs. K51t-TRIM37, ns P = 0.201, K51-TRIM37 vs. K51t-TRIM37, *P = 0.044, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (J) The xenograft tumor volume measurements of NSG mice injected with KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. For 7 days, K51t, *P = 0.018, K51-TRIM37, *P = 0.033, and K51t-TRIM37, ***P = 6.28×10^{-5}. For 10 days, K51t, *P = 0.011, K51-TRIM37, *P = 0.039, and K51t-TRIM37, **P = 0.007. For 13 days, K51t, ***P = 1.71×10^{-5}, K51-TRIM37, ***P = 0.0002, and K51t-TRIM37, ***P = 1.42×10^{-5}. For 16 days, K51t, **P = 0.002, K51-TRIM37, **P = 0.005, and K51t-TRIM37, ***P = 1.85×10^{-5}, K51t vs. K51t-TRIM37, **P = 0.005, and K51-TRIM37 vs. K51t-TRIM37, **P = 0.004, unpaired t test. Data are mean ± SD of biological replicates, n = 6/group. (K, L) Representative images of H&E (K, ×20, scale bar, 200 µm) and Ki67 (L, ×20, scale bar, 200 µm) for the tumors isolated in (J) are shown. (M–O) Representative immunohistochemical (IHC) images of ER (M, ×20, scale bar, 200 µm), PR (N, ×20, scale bar, 200 µm), and Her2 (O, ×20, scale bar, 200 µm) for the tumors isolated in (J) are shown. The manual quantification of the IHC signal is shown for ER (M, bottom), PR (N, bottom), and Her2 (O, bottom) (n = ~1200–1500 cells per group).
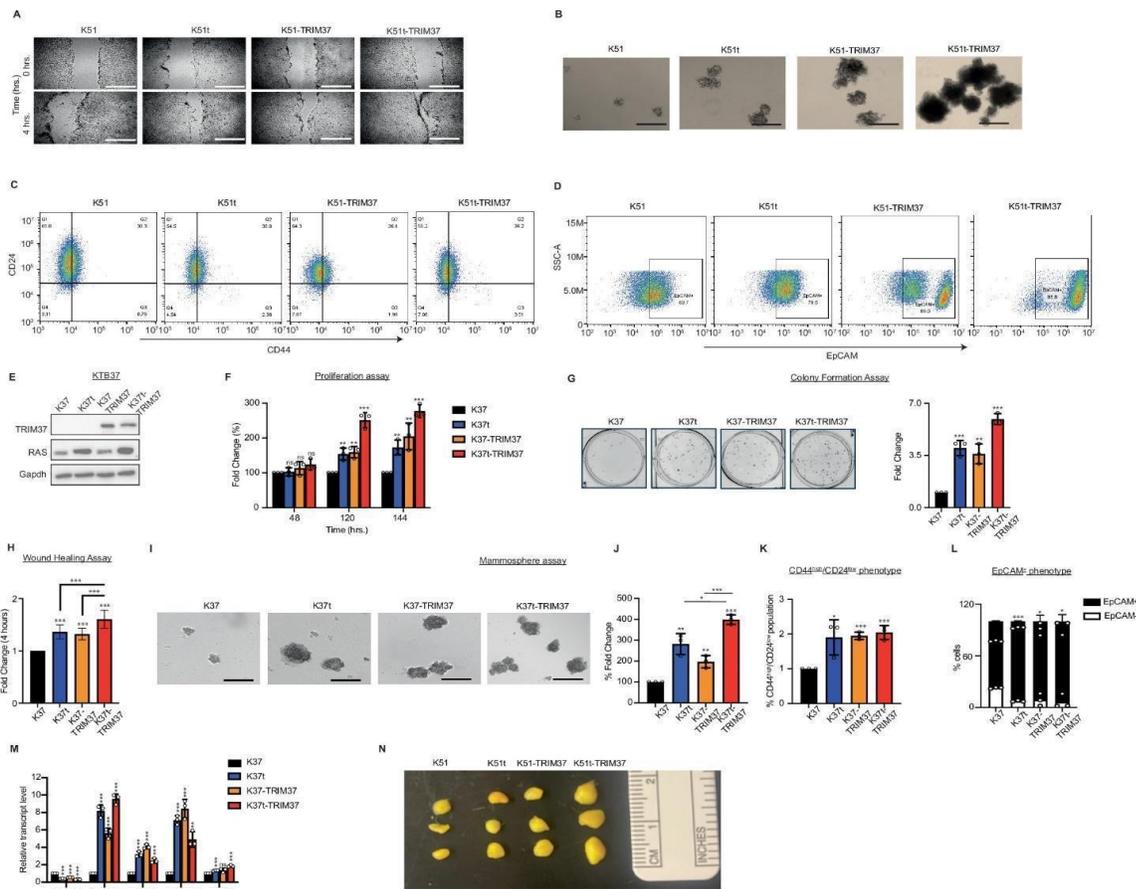
Supplementary Figure 4-5. *TRIM37*-directs neoplastic transformations in immortalized epithelial cells. (A) Representative phase contrast images of relative migratory abilities for KTB51, K51t, K51-TRIM37, and K51t-TRIM37 after 4 h. (10X, scale bar, 300 μm). (B) Representative images of solid mammospheres formed by KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. (10X, scale bar, 300 μm). (C, D) Representative FACS plot showing gating strategy and distribution of stained population for CD24 and CD44 (C) and EpCAM (D) in KTB51, K51t, K51-TRIM37, and K51t-TRIM37 cells. (E) Immunoblots in KTB37 and TRIM37 derivatives of RAS-transformed KTB37(K37t) cells. Gapdh was the loading control. (F) Relative cell growth for KTB37 and TRIM37 derivatives of RAS-transformed KTB37 (K37t-TRIM37) cells at indicated times. For 48 h., K37t, ns P = 0.655, K37-TRIM37, ns P = 0.287, and K37t-TRIM37, ns P = 0.058. For 120 h., K37t, **P = 0.006, K37-TRIM37, **P = 0.003, and K37t-TRIM37, ***P = 0.0003. For 144 h. K37t, **P = 0.004, K37-TRIM37, **P = 0.009, and K37t-TRIM37, ***P = 0.0002, unpaired t test. Data are mean ± SD of biological replicates, n = 3/group. (G) Representative bright-field images after crystal violet staining show the growth of KTB37, K37t, K37-TRIM37, and K37t-TRIM37 cells. The colonies were quantified (Right). For K37t, ***P = 0.0005, K37-TRIM37, **P = 0.003, and K37t-TRIM37, ***P = 1.86*10-5, unpaired t test. Data are mean ± SD of biological replicates, n = 3/group. (H) The relative migratory abilities for KTB37, K37t, K37-TRIM37, and K37t-TRIM37 cells were quantitated after 4 h. For K37t, ***P = 1.32*10-13, K37-TRIM37, ***P = 3.39*10-14, and K37t-TRIM37, ***P = 1.72*10-16; K37t vs. K37t-TRIM37,

***P = 4.13*10-5, K37-TRIM37 vs. K37t-TRIM37, ***P = 1.28*10-6, unpaired t test. Data are mean±SD of biological replicates, n = 6/group. (I, J) Representative images (10X, scale bar, 300 μm) (I) and quantitation (J) of solid mammospheres formed by KTB37, K37t, K37-TRIM37, and K37t-TRIM37 cells. The colonies were quantified. For K37t, **P = 0.003, K37-TRIM37, **P = 0.004, and K37t-TRIM37, ***P = 2.15*10-5; K37t vs. K37t-TRIM37, *P = 0.021, K37-TRIM37 vs. K37t-TRIM37, ***P = 0.0007, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (K, L) FACS analysis of CD24 and CD44 (K) and EpCAM (L) in KTB37, K37t, K37-TRIM37, and K37t-TRIM37 cells derived from mammospheres in (I, J). For CD44high/CD24low in K37t, *P = 0.036, K37-TRIM37, ***P = 0.0001, and K37t-TRIM37, ***P = 0.0008. For EpCAM+ in K37t, ***P = 1.99*10-5, K37-TRIM37, *P = 0.028, and K37t-TRIM37, *P = 0.024, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (M) qRT-PCR analysis of EMT markers in KTB37, K37t, K37-TRIM37, and K37t-TRIM37 cells derived from mammospheres in (I, J). Gapdh is used as an endogenous control. For CDH1 in K37t, ***P = 2.03*10-7, K37-TRIM37, ***P = 1.01*10-6, and K37t-TRIM37, ***P = 1.58*10-5. For CDH2 in K37t, ***P = 5.52*10-5, K37-TRIM37, ***P = 0.0001, and K37t-TRIM37, ***P = 1.38*10-5. For SNAI1 in K37t, ***P = 0.0002, K37-TRIM37, ***P = 2.56*10-5, and K37t-TRIM37, ***P = 0.0006. For TWIST1 in K37t, ***P = 3.08*10-5, K37-TRIM37, ***P = 0.0003, and K37t-TRIM37, **P = 0.002. For ZEB1 in K37t, ***P = 0.0001, K37-TRIM37, ns P = 0.051, and K37t-TRIM37, ***P = 0.0002, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (N) Representative bright-field images showing gross histology of xenografts harvested from NSG mice injected with, K51t, K51-TRIM37, and K51t-TRIM37 cells.

Context-dependent impact of oncogenic TRIM37 expression on breast cancer onset and progression

Tumorigenesis is a multistage process that progresses through accumulating numerous somatic mutations, epigenetic aberrations, and transcriptional alterations. To experimentally address the differential effect of TRIM37 on transformed phenotype in a cellular context, we established a reversible system with doxycycline-inducible TRIM37 in constitutively KTB51t cells (K51ti; Supplementary Fig.4-6A). In K51ti cells, adding doxycycline leads to a reversible expression of TRIM37 (Fig. 4-6A), and increased growth as measured through proliferation (Supplementary Fig.4-6B), and colony formation assay (Supplementary Fig.4-6C). Next, we examined the impact of TRIM37 depletion on

the loss of identity and cellular reprogramming phenotype in K51ti cells. As

expected, TRIM37 expression in the presence of doxycycline significantly

increased the number of mammospheres formed compared with the control KTB51

and uninduced K51ti cells (Figs. 4-6B and Supplementary Fig.4-6D). Interestingly,

doxycycline withdrawal stalled the reprogramming phenotype, and fewer or smaller

mammospheres were observed (Figs. 4-6B and Supplementary Fig.4-6D). In

concert with the mammospheres assay, the FACS analysis confirmed a dramatic

decrease in the CD44high/CD24low and EpCAM-positive cell population derived

from mammospheres following doxycycline withdrawal (Figs. 4-6C,D and

Supplementary Fig.4-6E,F).

To explore this phenomenon *in vivo*, we assessed the growth of K51ti

xenografts in NSG mice. Injected mice were fed on either a control or doxycycline-

containing diet to allow for the continuous expression of TRIM37, and tumor

growth was monitored. All the animals fed a doxycycline diet showed continued

tumor growth, whereas 100% of the control animals failed to show growth of

xenograft tumors (Fig. 4-6E). Interestingly, a subset of tumor-bearing animals,

when deprived of a doxycycline diet, showed a marked reduction in the tumor

growth, indicating the role of TRIM37 in early stages of tumorigenesis (Fig. 4-6E).

H&E and Ki67 staining revealed gross histology and the hyperproliferative cells of

tumors (Fig. 4-6F,G). The K51ti xenografts from mice fed on a doxycycline diet

demonstrated the absence of ER, PR, and Her2 expression relative to positive

control (Fig. 4-6H–J). Together, these results indicate that TRIM37 upregulation in

the early stages of the disease gives the tumor a "head-start" to progress and

metastasize by triggering cellular reprogramming.
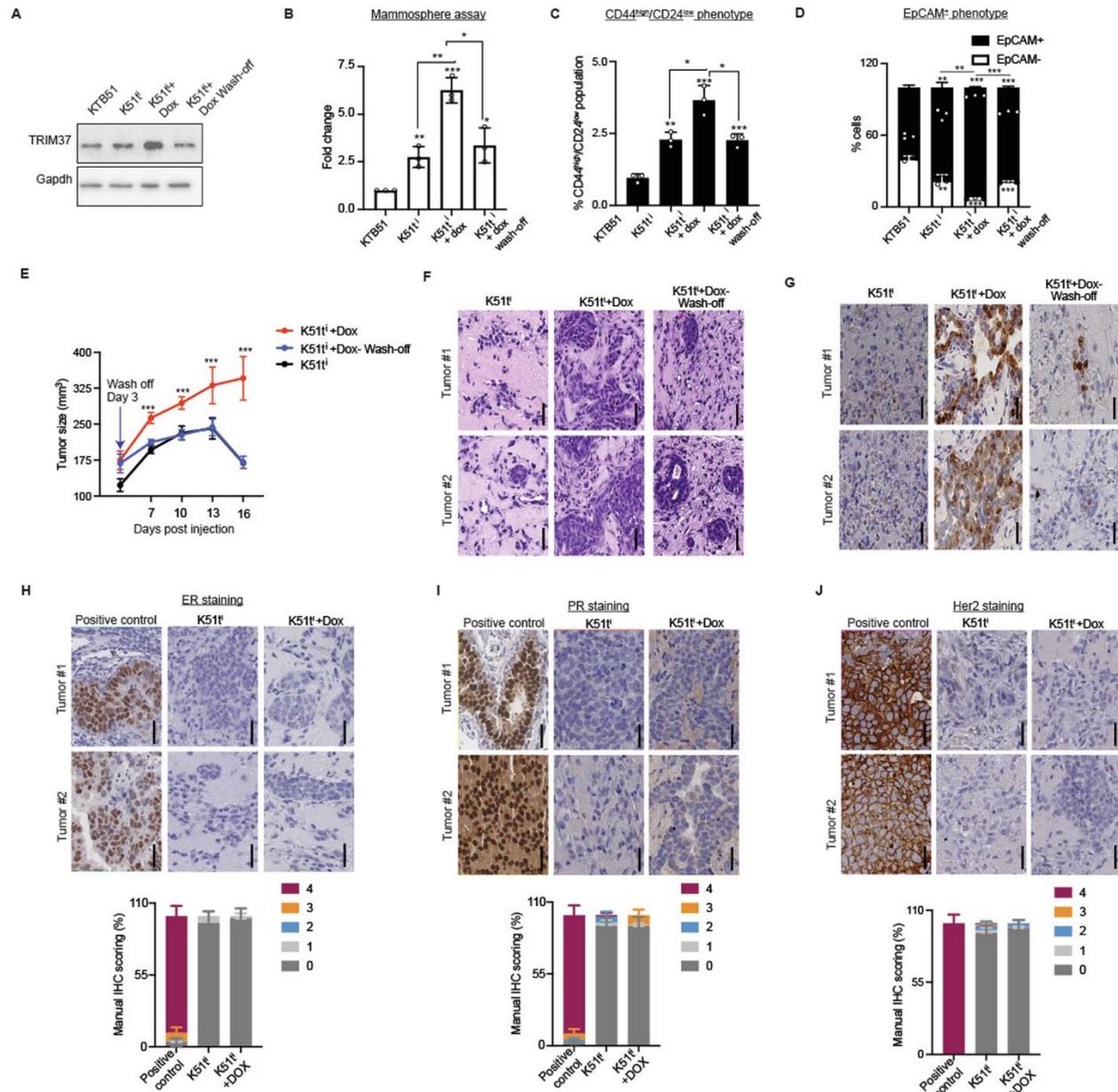
Figure 4-6. Early-stage expression of *TRIM37* affects breast cancer initiation and progression. (A) Immunoblot analysis in K51ti cells following treatment with doxycycline (dox) 48 h. post-induction or removal after 3 days (Wash-off). Gapdh was the loading control. (B)

Quantitation of solid mammospheres formed by K51ti in the presence of dox. For a subset of induced K51ti cells, dox was removed after 3 days of treatment (wash-off). For K51ti, **P = 0.005, K51ti+dox, ***P = 0.0002, K51ti+wash-off, *P = 0.011, K51ti vs. K51ti+dox, **P = 0.002, and K51ti+Dox vs. K51ti+wash-off, *P = 0.012, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (C, D) FACS analysis of K51ti mammospheres derived in (B) for CD24 and CD44 (C) and EpCAM (D) in the presence of dox or wash-off. For CD44high/CD24low in K51ti, **P = 0.001, K51ti+dox, ***P = 0.0009, K51ti+wash-off, ***P = 0.00098, K51ti vs. K51ti+dox, *P = 0.014, K51ti+dox vs. K51ti+wash-off, P = 0.012. For EpCAM+ in K51ti, **P = 0.003, K51ti+dox, ***P = 1.2×10^{-5}, K51ti+wash-off, ***P = 0.0001, K51ti vs. K51ti+dox, **P = 0.004, K51ti+dox vs. K51ti+wash-off, ***P = 6.65×10^{-5}, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (E) The xenograft tumor volume measurements of NSG mice injected with K51ti and fed on either a control or dox-enriched diet. A subset of animals was deprived of dox induction after 4 days (Wash-off). For 7 days, K51ti vs. K51ti+dox, ***P = 3.74×10^{-7}, and K51ti+dox vs. K51ti+wash-off, ***P = 2.22×10^{-6}. For 10 days, K51ti vs. K51ti+dox, ***P = 1.28×10^{-5}, and K51ti+dox vs. K51ti+wash-off, ***P = 2.16×10^{-6}. For 13 days, K51ti vs. K51ti+dox, ***P = 9.4×10^{-5}, and K51ti+dox vs. K51ti+wash-off, ***P = 4.31×10^{-5}. For 16 days, K51ti vs. K51ti+dox, ***P = 1.22×10^{-7}, and K51ti+dox vs. K51ti+wash-off, ***P = 3.74×10^{-7}, unpaired t test. Data are mean ± SD of biological replicates, n = 6/group. (F, G) Representative images of H&E (F, ×20, scale bar, 200 µm) and Ki67 (G, ×20, scale bar, 200 µm) for the tumors isolated in (E) are shown. (H–J) Representative immunohistochemical (IHC) images of ER (H, ×20, scale bar, 200 µm), PR (I, ×20, scale bar, 200 µm), and Her2 (J, ×20, scale bar, 200 µm) for the tumors isolated in (E) are shown. The manual quantification of the IHC signal is shown for ER (H, bottom), PR (I, bottom), and Her2 (J, bottom) (n = ~1200 cells per group.
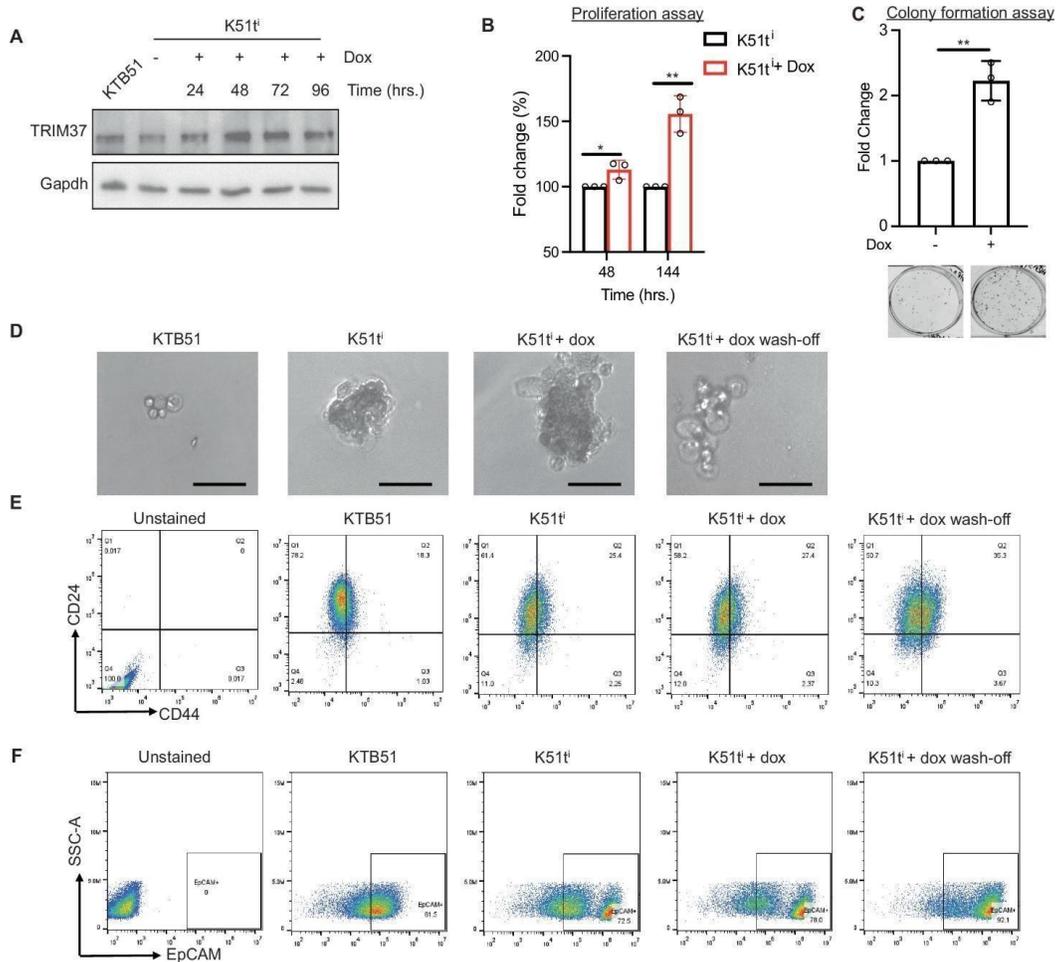
Supplementary Figure 4-6. TRIM37 is required for breast cancer onset. (A) Immunoblot analysis in K51ti cells following doxycycline (dox) treatment for 24, 48, 72 and 96 h. post-induction. Gapdh was the loading control. (B) Relative cell growth for K51ti and dox-induced K51ti at indicated times. For 48 h, *P = 0.034 and 144 h, **P = 0.002, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (C) The colony formation assay quantitating the growth of K51ti and dox-induced K51ti. Representative bright-field images after crystal violet staining are shown (Bottom). For K51ti vs. K51ti+dox, **P = 0.003, unpaired t test. Data are mean±SD of biological replicates, n = 3/group. (D) Representative images of solid mammospheres formed by K51ti and dox-induced K51ti at indicated times. (×10, scale bar, 300 μm). (E, F) Representative FACS plot showing gating strategy and distribution of stained population for CD24 and CD44 (E) and EpCAM (F) in K51ti and dox-induced K51ti at indicated times.

4.5 Discussion

The TNBC phenotype accounts for 30% of all breast cancer cases in BW [11], whereas only 15% of WW are diagnosed with TNBC [45]. Likewise, the

TNBC mortality rate is 40% higher in BW than in WW [46]. The efforts have been focused on understanding the molecular basis of racial disparity and characterizing the effects of African ancestry on breast cancer onset, progression, and clinical outcome. Significant genetic, epigenetic, and transcriptomic diversity has been identified through high-throughput genomics, proteomics, and cross-sectional imaging analysis of breast tumors from self-reported BW and WW [13,18,47,48]. For example, the mutation rate and intra-tumoral heterogeneity in breast tumors are significantly higher in BW compared to WW [49, 50]. The global transcript studies have detailed over 400 DEGs between the two racial identities [13, 18]. While these results are promising, there is a significant gap in our understanding of how these changes in gene expression are associated with cancer disparity and what regulates racial identity-specific gene expression. We note that the interpretation of the data is further complicated by the inherent limitations of small cohort size, with primary data from populations of European descent underscoring the need for better representations of diverse populations. Biases in analysis can also originate from the self-reported medical history data and heterogeneity of the racial identity. The lack of a diverse set of ancestry-specific cellular models and missing functional mechanisms linking genes or mutations to the unique TNBC biology in BW further present challenges in identifying and validating ancestry-specific genetic drivers that can be used as predictive biomarkers.

In this study, we used a candidate-based approach to interrogate the association of biological components to disproportionate TNBC incidence and mortality in BW. We performed the meta-analysis for gene-level associations to examine whether differences in the TRIM37 expression would account for the more aggressive TNBC phenotype in BW. We show that TRIM37 is expressed significantly higher in the cancer-free breast tissue from BW than in WW. We note that racial identity is a complex construct, and the use of self-reported racial identity is problematic for determining underlying ancestry-specific genetic determinants of breast cancer risk. We, therefore, estimated the global genomic ancestry for each individual in the TCGA cohort to evaluate the varying levels of admixture. Our analysis of the six samples from BW showed a high level of African ancestry (median 94.5%) with smaller contributions of European (median 5%) and Asian (median 0%) ancestry. Similarly, the WW samples showed a high level of European ancestry (median 99%). The samples from WW had a mixture of African (median 0.95%) and Asian ancestry (median 1.77%), with one sample showing greater than 30% Asian ancestry. Given that ancestral differences can impact the gene-disease associations, we believe the admixture refinement strongly supports TRIM37 function as a genetic determinant of racial disparity in TNBC patients.

Multiple SNPs have been associated with breast cancer prevalence, as well as clinicopathologic status [51-53]. Besides BRCA1 and BRCA2 mutations that

markedly increase breast cancer risk [54], hundreds of low- and moderate-risk susceptibility variants have been identified, including caspase-8 (rs2293554, rs6723097, [55]), TIMP-2 (rs7501477 [56]), and FSCN1 (rs56156320, rs3801004, [57]). In addition, SNPs play a critical role in phenotypic variation and disease susceptibility through effects on gene expression, such as PDE4DIP (rs12124527, [58]) and FTO (rs9939609, [59]). Interestingly, polymorphic variants in the tripartite motif-containing (TRIM) superfamily members have been linked to several human diseases, highlighting their clinical significance. For example, a recent study identified the breast cancer risk locus SNP rs4971059 that regulates the TRIM46-HDAC1 axis to modulate genes involved in DNA replication and repair [60]. Notably, TRIM37 polymorphic variants have previously been associated with leukemia (rs11656413, [61]), schizophrenia ((rs4968363, rs2877926, [62]), and inflammatory bowel [63].

Previous GWAS studies have identified multiple breast cancer susceptibility variants [64, 65], but ancestry-specific variants associated with breast cancer progression and aggressiveness are rare. Here, we identified a risk variant of rs57141087 in the 17q23 region that associates with TRIM37 expression through modulating promoter–enhancer interactions. We note that rs57141087 has not been previously identified as a causal variant for breast cancer, nor is it in strong LD with previously described risk SNPs. However, the discovery of all risk-associated

variants cannot be based on statistical analyses [66]. Furthermore, the underrepresentation of African ancestry and predominance of data collected from populations of European ancestry biases analyses and interpretation of the results. Significantly, our analysis of GWAS data from samples of African ancestry (Supplementary Fig.4-3D), complemented by comprehensive functional analysis, showed a significant association of rs57141087 with TNBC risk.

Our sequence analysis identified a unique site for NRF1 in the enhancer region of TRIM37 harboring risk variant rs57141087. Mechanistically, NRF1 is a transcription factor that functions as a homodimer to adopt a U-turn conformation, facilitating the interaction between the two TGCGC motifs in the GCGCATGCGC consensus sequence [43]. NRF1 activity is significantly associated with breast cancer and patient survival [67]. Interestingly, the transcription factor target enrichment analysis in the TCGA cohort identified increased NRF1 activity in BW with TNBC subtypes [68]. We found that NRF1 bound to risk variant rs57141087 modulates TRIM37 promoter activity in an allele-specific manner. Here, we show that the additional NRF1 binding in the enhancer region facilitates promoter–enhancer interaction through looping to increase TRIM37 promoter activity. Despite making progress in early diagnosis, optimized management, and breast cancer treatment, our understanding of cancer causation is limited. Detecting driver events in cancer initiation is pivotal for developing new predictive prognostic

biomarkers. We find that TRIM37 could contribute to disease susceptibility through its role in the early phase of carcinogenesis. Here, we uncovered that higher TRIM37 expression in the normal, cancer-free breast tissue of BW favors transcriptional activation sufficient to induce transformation in immortalized breast epithelial cells.

4.6 References

1.   Siegel RL, Miller KD, Wagle NS, Jemal A (2023) Cancer statistics, 2023. CA Cancer J Clin 73: 17-48

2.   Almansour NM (2022) Triple-Negative Breast Cancer: A Brief Review About Epidemiology, Risk Factors, Signaling Pathways, Treatment and Role of Artificial Intelligence. Front Mol Biosci 9: 836417

3.   Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 13: 4429-4434

4.   Jiagge E, Jibril AS, Chitale D, Bensenhaver JM, Awuah B, Hoenerhoff M, Adjei E, Bekele M, Abebe E, Nathanson SD et al (2016) Comparative Analysis of Breast Cancer Phenotypes in African American, White American, and West Versus East African patients: Correlation Between African Ancestry and Triple-Negative Breast Cancer. Ann Surg Oncol 23: 3843-3849

5.   Martini R, Newman L, Davis M (2022b) Breast cancer disparities in outcomes; unmasking biological determinants associated with racial and genetic diversity. Clin Exp Metastasis 39: 7-14

6.   Newman LA, Jenkins B, Chen Y, Oppong JK, Adjei E, Jibril AS, Hoda S, Cheng E, Chitale D, Bensenhaver JM et al (2019) Hereditary Susceptibility for Triple Negative Breast Cancer Associated With Western Sub-Saharan African

Ancestry: Results From an International Surgical Breast Cancer Collaborative. Ann Surg 270: 484-492

7.   Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, Schwartz GF, Park PK, Rosenberg AL, Brill K et al (2007) Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 110: 876-884

8.   Rapiti E, Pinaud K, Chappuis PO, Viassolo V, Ayme A, Neyroud-Caspar I, Usel M, Bouchardy C (2017) Opportunities for improving triple-negative breast cancer outcomes: results of a population-based study. Cancer Med 6: 526-536

9.   Garlapati C, Joshi S, Sahoo B, Kapoor S, Aneja R (2019) The persisting puzzle of racial disparity in triple negative breast cancer: looking through a new lens. Front Biosci (Schol Ed) 11: 75-88

10.   Newman LA, Kaljee LM (2017) Health Disparities and Triple-Negative Breast Cancer in African American Women: A Review. JAMA Surg 152: 485-493

11.   Siddharth S, Sharma D (2018) Racial Disparity and Triple-Negative Breast Cancer in African-American Women: A Multifaceted Affair between Obesity, Biology, and Socioeconomic Determinants. Cancers (Basel) 10

12.   Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, Wang X, Ademuyiwa F, Ahmed S, Ambrosone CB et al (2011) A common variant

at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet 43: 1210-1214

13.   Martini R, Delpe P, Chu TR, Arora K, Lord B, Verma A, Bedi D, Karanam B, Elhussin I, Chen Y et al (2022a) African Ancestry-Associated Gene Expression Profiles in Triple-Negative Breast Cancer Underlie Altered Tumor Biology and Clinical Outcome in Women of African Descent. Cancer Discov 12: 2530-2551

14.   Parada H, Jr., Sun X, Fleming JM, Williams-DeVane CR, Kirk EL, Olsson LT, Perou CM, Olshan AF, Troester MA (2017) Race-associated biological differences among luminal A and basal-like breast cancers in the Carolina Breast Cancer Study. Breast Cancer Res 19: 131

15.   Siddharth S, Parida S, Muniraj N, Hercules S, Lim D, Nagalingam A, Wang C, Gyorffy B, Daniel JM, Sharma D (2021) Concomitant activation of GLI1 and Notch1 contributes to racial disparity of human triple negative breast cancer progression. Elife 10

16.   Rahman, N., Realizing the promise of cancer predisposition genes. Nature, 2014. 505(7483): p. 302-8.

17.   Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C, Wahlstrom C, Pooley KA, Parsons MT, Fortuno C et al (2021) Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med 384: 428-439

18.   Marino N, German R, Podicheti R, Rusch DB, Rockey P, Huang J, Sandusky GE, Temm CJ, Althouse S, Nephew KP et al (2022) Aberrant epigenetic and transcriptional events associated with breast cancer risk. Clin Epigenetics 14: 21

19.   Bhatnagar S, Gazin C, Chamberlain L, Ou J, Zhu X, Tushir JS, Virbasius CM, Lin L, Zhu LJ, Wajapeyee N et al (2014) TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. Nature 516: 116-120

20.   Bhatnagar S, Green MR (2015) TRIMming down tumor suppressors in breast cancer. Cell Cycle 14: 1345-1346

21.   Przanowski P, Lou S, Tihagam RD, Mondal T, Conlan C, Shivange G, Saltani I, Singh C, Xing K, Morris BB et al (2020) Oncogenic TRIM37 Links Chemoresistance and Metastatic Fate in Triple-Negative Breast Cancer. Cancer Res 80: 4791-4804

22.   Przanowski P, Wasko U, Zheng Z, Yu J, Sherman R, Zhu LJ, McConnell MJ, Tushir-Singh J, Green MR, Bhatnagar S (2018) Pharmacological reactivation of inactive X-linked Mecp2 in cerebral cortical neurons of living mice. Proc Natl Acad Sci U S A 115: 7991-7996

23.   Kumar B, Prasad M, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, Storniolo AM, Rao X, Liu S, Wan J et al (2018) Normal Breast-Derived Epithelial Cells with Luminal and Intrinsic Subtype-Enriched Gene Expression Document

Interindividual Differences in Their Differentiation Cascade. Cancer Res 78: 5107-5123

24. Thomsen C, Nielsen S, Nielsen BS, Pedersen SH, Vyberg M (2020) Estrogen Receptor-alpha Quantification in Breast Cancer: Concordance Between Immunohistochemical Assays and mRNA-In Situ Hybridization for ESR1 Gene. Appl Immunohistochem Mol Morphol 28: 347-353

25. Panigrahi AK, Foulds CE, Lanz RB, Hamilton RA, Yi P, Lonard DM, Tsai MJ, Tsai SY, O'Malley BW (2018) SRC-3 Coactivator Governs Dynamic Estrogen-Induced Chromatin Looping Interactions during Transcription. Mol Cell 70: 679-694 e677

26. Rebouissou C, Sallis S, Forne T (2022) Quantitative Chromosome Conformation Capture (3C-qPCR). Methods Mol Biol 2532: 3-13

27. Hua JT, Ahmed M, Guo H, Zhang Y, Chen S, Soares F, Lu J, Zhou S, Wang M, Li H et al (2018) Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. Cell 174: 564-575 e518

28. Tihagam RD, Bhatnagar S (2023) A multi-platform normalization method for meta-analysis of gene expression data. Methods 217: 43-48

29. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655-1664

30.   Gyorffy B (2024) Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors. Innovation (Camb) 5: 100625

31.   Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31: 3555-3557

32.   Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet 103: 338-348

33.   Lowy-Gallego E, Fairley S, Zheng-Bradley X, Ruffier M, Clarke L, Flicek P, Genomes Project C (2019) Variant calling on the GRCh38 assembly with the data from phase three of the 1000 Genomes Project. Wellcome Open Res 4: 50

34.   Jia G, Ping J, Guo X, Yang Y, Tao R, Li B, Ambs S, Barnard ME, Chen Y, Garcia-Closas M et al (2024) Genome-wide association analyses of breast cancer in women of African ancestry identify new susceptibility loci and improve risk prediction. Nat Genet 56: 819-826

35.   Saleh M, Chandrashekar DS, Shahin S, Agarwal S, Kim HG, Behring M, Shaikh AJ, Moloo Z, Eltoum IA, Yates C, Varambally S, Manne U (2021) Gene Expression Omnibus GSE142731

36.   Denker A, de Laat W (2016) The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 30: 1357-1382

37.   Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503: 290-294

38.   Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12: 283-293

39.   Spurrell CH, Dickel DE, Visel A (2016) The Ties That Bind: Mapping the Dynamic Enhancer-Promoter Interactome. Cell 167: 1163-1166

40.   Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E et al (2013) Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 504: 306-310

41.   Rathinam C, Geffers R, Yucel R, Buer J, Welte K, Moroy T, Klein C (2005) The transcriptional repressor Gfi1 controls STAT3-dependent dendritic cell development and function. Immunity 22: 717-728

42.   Jensen KC, Higgins JP, Montgomery K, Kaygusuz G, van de Rijn M, Natkunam Y (2007) The utility of PAX5 immunohistochemistry in the diagnosis of undifferentiated malignant neoplasms. Mod Pathol 20: 871-877

43.   Liu K, Li W, Xiao Y, Lei M, Zhang M, Min J (2024) Molecular mechanism of specific DNA sequence recognition by NRF1. Nucleic Acids Res 52: 953-966

44.   Virbasius CA, Virbasius JV, Scarpulla RC (1993) NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. Genes Dev 7: 2431-2445

45.   Bowen RL, Duffy SW, Ryan DA, Hart IR, Jones JL (2008) Early onset of breast cancer in a group of British black women. Br J Cancer 98: 277-281

46.   Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer Statistics, 2021. CA Cancer J Clin 71: 7-33

47.   Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, Yoshimatsu TF, Pitt JJ, Hoadley KA, Troester M et al (2017) Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. JAMA Oncol 3: 1654-1662

48.   Lee CI, Zhu W, Onega T, Henderson LM, Kerlikowske K, Sprague BL, Rauscher GH, O'Meara ES, Tosteson ANA, Haas JS et al (2021) Comparative Access to and Use of Digital Breast Tomosynthesis Screening by Women's Race/Ethnicity and Socioeconomic Status. JAMA Netw Open 4: e2037546

49.   Ansari-Pour N, Zheng Y, Yoshimatsu TF, Sanni A, Ajani M, Reynier JB, Tapinos A, Pitt JJ, Dentro S, Woodard A et al (2021) Whole-genome analysis of Nigerian patients with breast cancer reveals ethnic-driven somatic evolution and distinct genomic subtypes. Nat Commun 12: 6946

50.   Parida S, Siddharth S, Xia Y, Sharma D (2023) Concomitant analyses of intratumoral microbiota and genomic features reveal distinct racial differences in breast cancer. NPJ Breast Cancer 9: 4

51.   Middha P, Wang X, Behrens S, Bolla MK, Wang Q, Dennis J, Michailidou K, Ahearn TU, Andrulis IL, Anton-Culver H et al (2023) A genome-wide gene-environment interaction study of breast cancer risk for women of European ancestry. Breast Cancer Res 25: 93

52.   Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ (2018) Common Genetic Variation and Breast Cancer Risk-Past, Present, and Future. Cancer Epidemiol Biomarkers Prev 27: 380-394

53.   Stolarova L, Kleiblova P, Zemankova P, Stastna B, Janatova M, Soukupova J, Achatz MI, Ambrosone C, Apostolou P, Arun BK et al (2023) ENIGMA CHEK2gether Project: A Comprehensive Study Identifies Functionally Impaired CHEK2 Germline Missense Variants Associated with Increased Breast Cancer Risk. Clin Cancer Res 29: 3037-3050

54.   Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T et al (2010) A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat Genet 42: 885-892

55.   Park HL, Ziogas A, Chang J, Desai B, Bessonova L, Garner C, Lee E, Neuhausen SL, Wang SS, Ma H et al (2016) Novel polymorphisms in caspase-8 are associated with breast cancer risk in the California Teachers Study. BMC Cancer 16: 14

56.   Peterson NB, Beeghly-Fadiel A, Gao YT, Long J, Cai Q, Shu XO, Zheng W (2009) Polymorphisms in tissue inhibitors of metalloproteinases-2 and -3 and breast cancer susceptibility and survival. Int J Cancer 125: 844-850

57.   Wang CQ, Tang CH, Wang Y, Jin L, Wang Q, Li X, Hu GN, Huang BF, Zhao YM, Su CM (2017) FSCN1 gene polymorphisms: biomarkers for the development and progression of breast cancer. Sci Rep 7: 15887

58.   Kim S, Cho H, Lee D, Webster MJ (2012) Association between SNPs and gene expression in multiple regions of the human brain. Transl Psychiatry 2: e113

59.   Karra E, O'Daly OG, Choudhury AI, Yousseif A, Millership S, Neary MT, Scott WR, Chandarana K, Manning S, Hess ME et al (2013) A link between FTO, ghrelin, and impaired brain food-cue responsivity. J Clin Invest 123: 3539-3551

60.   Zhang Z, Liu X, Li L, Yang Y, Yang J, Wang Y, Wu J, Wu X, Shan L, Pei F et al (2021) SNP rs4971059 predisposes to breast carcinogenesis and chemoresistance via TRIM46-mediated HDAC1 degradation. EMBO J 40: e107974

61.   Hess M, Lenz S, Blatte TJ, Bullinger L, Binder H (2017) Partitioned learning of deep Boltzmann machines for SNP data. Bioinformatics 33: 3173-3180

62.   Hall LS, Medway CW, Pain O, Pardinas AF, Rees EG, Escott-Price V, Pocklington A, Bray NJ, Holmans PA, Walters JTR et al (2020) A transcriptome-wide association study implicates specific pre- and post-synaptic abnormalities in schizophrenia. Hum Mol Genet 29: 159-167

63.   Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N et al (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43: 1066-1073

64.   Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087-1093

65.   Ferreira MA, Gamazon ER, Al-Ejeh F, Aittomaki K, Andrulis IL, Anton-Culver H, Arason A, Arndt V, Aronson KJ, Arun BK et al (2019) Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat Commun 10: 1741

66.   van de Bunt M, Cortes A, Consortium I, Brown MA, Morris AP, McCarthy MI (2015) Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. PLoS Genet 11: e1005535

67.   Falco MM, Bleda M, Carbonell-Caballero J, Dopazo J (2016) The pan-

cancer pathological regulatory landscape. Sci Rep 6: 39709

68.   Ramos J, Yoo C, Felty Q, Gong Z, Liuzzi JP, Poppiti R, Thakur IS, Goel R,

Vaid AK, Komotar RJ et al (2020) Sensitivity to differential NRF1 gene signatures

contributes to breast cancer disparities. J Cancer Res Clin Oncol 146: 2777-2815

Chapter 5: Conclusion

5.1 Summary of Major Contributions

This thesis explored the multifaceted role of TRIM37 in the progression of TNBC and its potential as a therapeutic target and biomarker. We found that TRIM37 is critical for TNBC cells to survive chemotherapy-induced stress and adapt during metastasis. Additionally, our research highlights significant racial disparities in TNBC, with TRIM37 expression potentially contributing to differences in disease prevalence, aggressiveness, and treatment response across populations. Below is a summary of the key findings that illustrate the role of TRIM37 in TNBC.

Mechanistic Insights

1. Role in chemoresistance

Chemoresistance in TNBC is a complex process often driven by genetic and epigenetic changes [1], with TRIM37 identified as a key player in this phenomenon. TRIM37 interacts with DNA repair pathways essential for maintaining genome integrity under chemotherapy-induced stress. Specifically, TRIM37 catalyzes the mono-ubiquitination of H2A, a chromatin modification that accumulates at DNA

damage sites, particularly double-strand breaks (DSBs). The repair of DSBs is crucial for cancer cell survival during chemotherapy, and the efficiency of this repair process often correlates with treatment resistance. In TNBC, TRIM37 expression is significantly upregulated following chemotherapy, a process driven by ATM/E2F1/STAT signaling. Notably, pharmacological inhibition of ATM or its downstream components significantly reduces TRIM37 expression and increases DNA damage, emphasizing its contribution to chemoresistance. Additionally, higher levels of TRIM37 in TNBC tumors promote resistance by facilitating tumor cell survival and protecting against chemotherapy-induced cell death. These findings suggest that TRIM37 serves as both a biomarker of chemoresistance and a critical mediator of the aggressive nature of TNBC cells.

2. Role in metastasis

Metastasis remains one of the most challenging and lethal aspects of cancer progression, particularly in aggressive subtypes such as TNBC. In addition to its role in chemoresistance, TRIM37 also plays a crucial role in promoting metastasis in TNBC. We demonstrate, for the first time, that TRIM37 in association with, the polycomb complex, alters gene expression to induce a pro-metastatic transcriptional program. This program includes key processes involved in metastasis, such as focal

adhesion, epithelial-mesenchymal transition (EMT), glycolysis, hypoxia, JAK/STAT signaling, apical junctions, immune response, and inflammatory-related genes. By enforcing a pro-metastatic transcriptional program, TRIM37 promotes tumor spread, making it a critical target for therapeutic intervention aimed at preventing or limiting metastasis in TNBC patients.

3. Role in racial disparity

This thesis also explores the role of TRIM37 in the racial disparity observed in TNBC, particularly in BW, who experience higher prevalence, more aggressive disease, and worse prognosis compared to WW. Elevated TRIM37 expression in cancer-free breast tissue from BW suggests it may contribute to the more aggressive TNBC phenotype seen in this population group. Additionally, a reporter-based screen revealed that the risk variant of rs57141087, in the enhancer region of TRIM37, increases its expression through the transcription factor NRF1. This variant, more prevalent in individuals of African ancestry, enhances TRIM37 promoter activity and may contribute to both the higher susceptibility to and more rapid progression of TNBC in BW.

Therapeutic implications

TRIM37 drives aggressive TNBC biology by promoting resistance to chemotherapy and inducing a pro-metastatic transcriptional program. Given its critical role in these processes, TRIM37 offers promising therapeutic implications, not only as a potential treatment target but also as a crucial biomarker for the screening and management of TNBC.

1. Nanoparticles

One novel therapeutic strategy explored here involves the selective delivery of TRIM37-specific antisense oligonucleotides (TRIM37-ASO) using nanoparticle-based systems. These nanoparticles were functionalized with Farletuzumab, an anti-folate receptor 1 (FOLR1) antibody, to specifically target TNBC cells that exhibit elevated FOLR1 expression. This selective delivery mechanism was shown to significantly reduce TRIM37 expression, thereby inhibiting the pro-metastatic and chemoresistant functions of TRIM37. These results highlight the therapeutic potential of combining TRIM37 inhibition with targeted nanoparticle delivery systems to reduce metastatic spread and overcome chemotherapy resistance in TNBC.

2. Predictive biomarker

TRIM37 expression is significantly higher in cancer-free breast tissue from BW compared to WW, and this elevated expression is associated with early-stage carcinogenesis. Our results suggest that increased TRIM37 levels in normal breast tissue contribute to the transformation of breast epithelial cells, highlighting its role in the initiation and progression of the disease. Given that TNBC disproportionately affects BW with more aggressive disease and poorer outcomes, TRIM37 could serve as a valuable biomarker for early detection and predicting the likelihood of aggressive disease.

TRIM37's expression is also linked to both chemoresistance and metastatic potential in TNBC, making it an essential marker for predicting disease aggression, treatment response, and chemotherapy failure. High TRIM37 levels are particularly indicative of poor prognosis and increased risk of metastatic progression. By incorporating TRIM37 into clinical screenings and measuring its expression in tumor biopsies, clinicians could better stratify patients based on their risk of developing aggressive TNBC or experiencing chemotherapy resistance.

5.2 Future directions

Building upon the findings of this thesis, future research will focus on (1) investigating additional molecular contributors to racial disparities in TNBC beyond TRIM37, and (2) developing small molecule inhibitors as clinically viable therapies

targeting TRIM37. These directions aim to deepen our mechanistic understanding of TNBC progression and help address the gap between lab discoveries and real-world treatments, making precision medicine more accessible.

Future direction I: Investigating Racially Segregating miRNA Variants in TNBC

    a) Rationale

While this thesis established TRIM37 as a major driver of chemoresistance, metastasis, and racial disparity in TNBC, it is likely that other regulatory factors contribute to the disease's heterogeneity and differential outcomes. Among these, microRNAs (miRNAs) are powerful post-transcriptional regulators of gene expression, influencing genes involved in DNA repair, metastasis, and immune response, all hallmarks of aggressive TNBC [1]. Importantly, miRNA expression is modulated by germline genetic variants, including single nucleotide polymorphisms (SNPs), many of which differ in frequency between racial groups [2]. Given the role of TRIM37 in racial disparity and the known regulatory influence of miRNAs, we hypothesize that racially segregating miRNA-associated variants may represent additional key contributors to TNBC progression and disparity.

miRNAs can simultaneously regulate multiple genes, including oncogenes and tumor suppressors, making them powerful molecular switches in cancer. Identifying

170

racially segregating variants that impact miRNA expression will uncover upstream regulators that may explain race-associated differences in TRIM37 levels and other oncogenic pathways. Moreover, miRNAs themselves may serve as biomarkers or therapeutic targets. By focusing on genetic variants that influence miRNA expression, rather than individual genes, we can capture a broader regulatory landscape potentially involved in the aggressive TNBC phenotype seen in Black women.

b) Approach

To identify racially segregating genetic variants associated with miRNA expression, we will utilize an integrative multi-step approach:

1. eQTL Analysis: Publicly available GWAS and miRNA expression datasets from breast cancer patients of African and European ancestry will be leveraged to identify SNPs associated with differential miRNA expression across racial groups. This analysis will identify SNPs that segregate by race and have a functional impact on miRNAs expression.

2. Functional Validation of miRNA Variants: Following the identification of racially segregating miRNA-associated SNPs, a small-scale luciferase reporter assay will assess the effect of SNP alleles on miRNA promoter activity. For each identified SNP, a double-stranded oligonucleotides containing the risk and

reference alleles will be designed and cloned into a luciferase reporter vector to assess how each SNP impacts miRNA gene promoter activity. Only positively scoring SNPs will be used for downstream studies.

Positively scoring SNPs will undergo endogenous validation using CRISPR/Cas9-mediated knock-in in breast epithelial cell lines to determine their impact on miRNA expression. This approach will involve introducing each SNP into established immortalized breast cancer cell lines (KTB51 and KTB37) using CRISPR/Cas9 technology. For each SNP of interest, we will use single guide RNAs (sgRNAs) to introduce the risk allele into the cells, with control cells harboring the reference allele. To confirm the successful integration of the SNPs, we will sequence the edited regions, ensuring the correct knock-in of the variants. As a negative control, a non-targeting control sgRNA will be used to ensure that the observed effects are due to the targeted SNP and not off-target effects. Then, the expression of the targeted miRNAs will be measured by quantitative PCR (qPCR) and western blot to confirm whether the SNP affects miRNA levels. Additionally, functional studies will be done to identify how positively scoring SNPs affect miRNAs expression.

3. Phenotypic Impact: To understand how the identified racially segregating miRNA variants influence breast cancer progression, we will examine the impact

of the SNPs on cell behavior. This will include assays to assess cell proliferation, migration, and invasion. We will compare the results between cells carrying the risk allele versus the reference allele for each miRNA-associated SNP.

To confirm the *in vitro* findings, we will utilize xenograft models in immunocompromised mice to study the effects of between cells carrying the risk allele versus the reference allele on tumor growth and metastasis. Breast cancer cells carrying the risk and reference alleles for each miRNA-associated SNP will be injected into the mammary fat pad of mice, and tumor progression will be monitored. Upon reaching a specified size, tumors will be harvested and analyzed for miRNA expression and gene expression changes to validate our *in vitro* findings in an *in vivo* setting.

c) Impact

This study will expand the current understanding of how inherited genetic variation influences TNBC risk and progression across racial groups. By identifying miRNA-mediated regulatory networks that drive tumor aggressiveness, this research could unveil new biomarkers for early detection and uncover novel therapeutic targets. Ultimately, this will support the development of more racially inclusive, genetically informed interventions for TNBC.

Future direction II: Development of Small Molecule Inhibitors Targeting

TRIM37

a) **Rationale**

While targeting TRIM37 with antisense oligonucleotides (ASOs) has shown

promise in reducing its expression and inhibiting the metastatic progression of

TNBC in this thesis, several challenges remain for translating ASO-based therapies

into clinical settings. These challenges include issues related to their stability,

bioavailability, and potential off-target effects [3]. Furthermore, the delivery of

ASOs requires complex, often costly, nanoparticle-based systems, which may not

always be feasible or efficient in clinical settings [4]. Thus, there is a pressing need

to develop small molecule inhibitors (SMIs) that directly target TRIM37. Contrary

to ASOs, SMIs offer improved tissue penetration, enhanced pharmacokinetic

properties, and greater cost-effectiveness, owing to their simpler manufacturing

processes and broader delivery methods [5]. Given TRIM37's critical E3 ligase

function specifically its RING domain-mediated mono-ubiquitination of histone

H2A, targeting its enzymatic activity with small molecules may present a more

viable therapeutic strategy.

TRIM37 regulates two of the most lethal aspects of TNBC: chemoresistance

and metastasis. Therefore, disrupting its E3 ligase activity could disable both DNA

174

damage repair mechanisms and pro-metastatic transcriptional programs simultaneously. Additionally, small molecule inhibitors are more suitable for long-term treatment strategies and can be more easily integrated into existing clinical workflows compared to nanoparticle-delivered ASOs.

b) Approach

A systematic approach integrating in silico modeling, *in vitro* functional assays, and *in vivo* validation will be used to identify small molecule inhibitors of TRIM37.

1. In Silico Modeling and Virtual Screening: Using the available crystal structure of the TRIM37 RING domain, a virtual high-throughput screening (vHTS) of diverse chemical libraries will be performed to identify small molecules predicted to interact with the RING domain. Computational docking simulations will be used to assess the binding orientation and affinity of candidate compounds. Promising hits will undergo molecular dynamics simulations to evaluate the stability of the protein-ligand interactions under physiological conditions.

2. In Vitro TRIM37 Ubiquitination Assays: Following the identification of small molecule candidates from virtual screening, their ability to inhibit TRIM37's ubiquitin ligase activity will be evaluated using in vitro ubiquitination assays. Recombinant TRIM37 protein or its RING domain will be purified and used to

catalyze the mono-ubiquitination of histone H2A or a synthetic H2AK119 substrate in the presence of small molecule inhibitors. The reaction products will be analyzed by Western blotting to detect the presence of ubiquitinated H2A. In addition, we will use ubiquitin-sensitive luciferase-based reporter assays to monitor the inhibition of TRIM37-mediated gene activation by the small molecules, providing a quantitative measure of their efficacy.

3. Secondary Screening for Specificity and Toxicity: To validate the specificity of the identified small molecules, we will perform secondary screening in TRIM37-expressing breast cancer cell lines, including TNBC models. These assays will involve treating cells with the small molecules and assessing the effects on TRIM37-mediated gene expression, cell proliferation, migration, and apoptosis. This will be followed by comparing the effects in TRIM37-knockdown or TRIM37-null cell lines to ensure that observed effects are indeed due to inhibition of TRIM37 activity. Moreover, a cytotoxicity assay will be conducted to assess the potential off-target effects of the small molecules and prioritize candidates with low cytotoxicity and minimal off-target activity.

4. In Vivo Validation Using Xenograft Models: Xenograft models will assess the impact of candidate inhibitors on tumor growth, metastasis, and TRIM37-related

signaling. These studies will validate the therapeutic relevance of TRIM37 inhibition in a clinically relevant context.

c) Impact

This work will deliver the first targeted small molecule inhibitors of TRIM37, a protein previously considered undruggable. If successful, this research will provide a new therapeutic option for patients with TNBC, particularly those with high TRIM37-expressing tumors.

5.3 References

1. Sun G, Yan J, Noltner K, et al., SNPs in human miRNA genes affect biogenesis and function. RNA, 2009.15(9).

2. Bensen, Jeannette T et al., Association of germline microRNA SNPs in pre-miRNA flanking region and breast cancer risk and survival: the Carolina Breast Cancer Study. Cancer causes & control, 2013. 24(6)

3. Dhuri, K., et al., Antisense Oligonucleotides: An Emerging Area in Drug Discovery and Development. J Clin Med, 2020. 9(6).

4. Gagliardi, M. and A.T. Ashizawa, The Challenges and Strategies of Antisense Oligonucleotide Drug Delivery. Biomedicines, 2021. 9(4).

5. Beck, S., et al., Small molecules and their impact in drug discovery: A perspective on the occasion of the 125th anniversary of the Bayer Chemical Research Laboratory. Drug Discov Today, 2022. 27(8).