One of these things is not like the others: investigating the role of germ cell gene GTSF1 in carcinogenesis.

Amelia Martínez Villarreal

Division of Experimental Medicine

Faculty of Medicine and Health Sciences

McGill University, Montreal

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Montréal, Québec, Canada

November 2024

To my mom and my grandma,

thanks for always being the wind under my wings.

To my dad and my grandpa,

thanks for never leaving me alone.

Table of contents

Abstract

Carcinogenesis is a multi-step process, and each step might provide malignant cells with a selective advantage over their neighbor cells. Emerging data demonstrates reactivation of developmental programs can provide cancer cells with selective advantages. Interestingly, expression of genes belonging to these programs present great therapeutic potential. Considering these genes are normally expressed in immune privileged organs (i.e. the immune system does not develop tolerance to antigens in these organs), their reactivation in somatic cancer cells renders them, in theory, highly immunogenic. In other words, the reactivation of these programs in cancer cells may lead to the production of tumor-specific antigens.

Therefore, our lab has focused on understanding the role the reactivation of the developmental programs, cancer-germline antigens and transposable elements, have in carcinogenesis. Interestingly, the ectopic expression of Gametocyte Specific Factor 1 (*GTSF1*) mRNA in cutaneous T-cell lymphoma (CTCL) has been associated with advanced disease stages and a worse prognosis. Thus, my research focused on understanding the role GTSF1 has in carcinogenesis. There is no previous functional analysis of this gene in carcinogenesis.

Here, I developed different cell line models to investigate the role of GTSF1 in carcinogenesis. GTSF1 silencing in a lung cancer model led to modification of the number of transposition events. This suggests that in lung cancer, GTSF1 controls the reactivation of transposons. In contrast, GTSF1 silencing in CTCL models led to T cell activation and production of the cytokines interferon gamma (IFNγ) and tumor necrosis factor alpha (TNFα). In other words, GTSF1 silencing led to a partial shift towards the effector phenotype. This suggests that in CTCL, GTSF1 modifies the memory/effector phenotype of the malignant cells. Furthermore, I evaluated

GTSF1 expression in CTCL clinical samples highlighting its association with advanced disease stage and worse prognosis. This suggests the potential of GTSF1 as a prognosis biomarker.

Taken together, my results suggest that the role GTSF1 has in carcinogenesis is specific to the cancer type expressing it. In CTCL, the potential role of GTSF1 is through the modification of the memory/effector phenotype, impacting survival and prognosis of patients. Overall, my results support the hypothesis that the reactivation of developmental programs can provide cancer cells with selective advantages.

Abrégé

Le développement du cancer est un processus à plusieurs étapes qui peut donner aux cellules cancéreuses un avantage sélectif sur leurs voisines. Les recherches les plus récentes suggèrent que la réactivation des programmes de développement peut conférer aux cellules cancéreuses des avantages sélectifs. Ainsi, l'expression des gènes impliqués dans ces programmes représente une voie thérapeutique intéressante, notamment, parce que ces gènes, dans des conditions saines, sont exprimés uniquement dans des organes immunologiquement privilégiés. Le système immunitaire n'a donc pas développé de tolérance aux antigènes qui peuvent être trouvés dans ces organes. C'est pourquoi, la réactivation des programmes de développement dans les cellules cancéreuses rendrait celles-ci hautement immunogènes.

Notre laboratoire s'est concentré sur la compréhension du rôle que deux programmes de développement, les antigènes tumoraux des cellules germinales et les transposons, dans le développement du cancer. Basé sur l'identification de l'expression anormale du gène Gametocyte Specific Factor 1 (*GTSF1*) chez les patients atteints de lymphomes cutanés à cellules T est associée à un stade plus avancé de la maladie, ainsi qu'à un pronostic plus faible, mon travail de doctorat a porté sur la compréhension du rôle de GTSF1 dans le développement du cancer. Mon travail répond alors à un besoin de connaissance non comblé puisqu'il n'existe actuellement aucune recherche sur le rôle de GTSF1 dans le développement du cancer.

Dans mon travail de doctorat, j'ai développé différents modèles cellulaires pour étudier le rôle de GTSF1 dans le développement du cancer. Dans un modèle de cancer du poumon, la répression génique de GTSF1 a entraîné des changements dans le nombre de transpositions. Ainsi, dans le cancer du poumon, le rôle de GTSF1 pourrait être de contrôler la réactivation des

transposons. D'autre part, la répression génique de GTSF1 dans des modèles de lymphome cutané à cellules T a entraîné l'activation des cellules T et la production de cytokines interféron gamma (IFNγ) et facteur de nécrose tumorale alpha (TNFα). La répression de GTSF1 pourrait donc conduire à une modification du phénotype des cellules malignes en un phénotype de cellules T effectrices. Enfin, l'analyse de l'expression de GTSF1 dans des échantillons de patients a permis de confirmer l'association entre l'expression de ce gène et les stades avancés de la maladie, et un mauvais pronostic. GTSF1 pourrait donc être utilisé comme marqueur pour évaluer le pronostic des patients atteints de lymphomes cutanés à cellules T.

En résumé, les travaux de mon doctorat suggèrent que le rôle de GTSF1 dépend du type de cancer dans lequel il est exprimé. En effet, dans les lymphomes cutanés à cellules T, GTSF1 modifie le phénotype des cellules effectrices et de la mémoire, affectant ainsi la survie et le pronostic des patients. De plus, mon travail doctoral soutient l'hypothèse selon laquelle la réactivation des programmes de développement des cellules cancéreuses leur confère un avantage sélectif.

Acknowledgements

The work presented here summarizes six years of my life learning about science and how to do science. It has been a long and strenuous road, filled with ups and downs but I am grateful for the opportunities that have allowed me to become the scientist I am today. And because no road is travelled alone, there are plenty of people I would like to thank.

First, thanks to my supervisor Dr. Ivan Litvinov for the opportunity to pursue my PhD studies in his lab. Thanks for sharing with me the fascinating field of cutaneous T-cell lymphomas and for the opportunity to collaborate with my home country, Mexico. Despite the challenges of this collaboration, I am proud to say that we did our best effort.

I would also like to thank my thesis committee members Dr. Suhad Ali, Dr. Jean-Jacques Lebrun, Dr. Yann V. Charli Joseph and Dr. Denis Sasseville. Your support, guidance, expertise and the balanced background between molecular scientists and dermatologists of the committee, pushed me to learn from each of you and to give my best every year. A special thanks to Dr. Suhad Ali, my academic advisor, for your guidance and support navigating the difficult times of my PhD. Thanks for always having time for a small chat and for your advice and support across the years.

Thanks to all my colleagues in the Litvinov lab: Dr. Philippe Lefrançois, Dr. Pinxing Xie, Scott Gun, Sriraam Sivachandran, Daniel Josue Guerra Ordaz, François Lagacé, Augustin Barolet and all other people I had the opportunity to learn from. A gigantic thank you to my lab friends throughout the years in the Litvinov lab: Dr. Jennifer Gantchev, Brandon Ramchatesingh, Dr. Marine Lambert and Dr. Raman Gill; thanks for all those lunch times filled with laughter and sharing. A special thanks to Brandon for your support these last two years, with experiments, thoughtful discussions or just listening; you will finally be able to work now that I am not around

to talk endlessly. An the biggest thank you to Jenn, my first Canadian friend. Without your support, I would not be the scientist I am today. Thanks for making my life in Canada a joyful adventure, thanks for teaching me about winters and for receiving me with your family, thanks for being a role model and an inspiration, thanks for always trying to make this journey less rough for me. I will forever be grateful for your friendship.

Many thanks to all the other amazing people I have had the pleasure to meet throughout these six years: Julia Messina-Pacheco and the rest of the lunch crew, Mónica Lara, Paulina Garcia de Alba, Jalal Al-Rahbani and Jesus Molina. The fun times and thoughtful discussions shared with you always refilled my energy tank. Thanks to my friends back in Mexico: Mily Ferrat, Beto Cristerna and Fer Araico. Thanks, because no matter the time and distance, you have always believed in me, thanks for your friendship all these years. And a huge thank you to Javier Rodriguez. Thanks for welcoming me to Montreal, for giving me all the details about living in this city, for always being a listening ear and for all the laughs and food. Your support and understanding have really made a difference in my journey.

I will forever be grateful for the amazing, incredible and supportive family I am so lucky to have. Thanks to each member of the Villarreal Dau family, each of you in your own way has supported my dreams and my passion for science. Thanks Leila, María, Valeria, Juan, Carlos, Miguel, Marcos, Luz Elena, Luis Marcos, David and David. Thanks to my best friend and sister, Tatiana. I am the luckiest cause you are my sister. Your laugh and kind words will always be the reason I give my best. Thanks to my grandma, ma. You are one of the reasons I am who I am today. Thanks for always filling me with love, for your endless support, for always believing in me. The biggest thank you to my mom, this degree is yours too. Life has put us in complicated situations, but you have always faced everything with hard work and love. Thanks for your love

and for everything you have taught me, from the love for knowledge to always believing in myself. And thanks to my dad and my grandad, pa. Something inside me has always known that, despite life decided it was time for you to leave this world, I have never been alone.

And finally, thanks to my partner, my love, Daniel Jimenez. Thanks for being my calm and my peace. Thanks for always believing in me and for your patience. Thanks for always encouraging me to be the best version of myself in all aspects of life. This degree has been strongly intertwined with our journey and I will forever be thankful for your love.

Most importantly, thanks to little me, little Amelia. Your passion has taken you to incredible places.

Contributions to original knowledge

Chapter II – Materials and Methods

1. To understand the role of Gametocyte Specific Factor 1 (GTSF1) in carcinogenesis, I developed the following models:
    a. Cutaneous T-Cell Lymphoma (CTCL) cell lines with GTSF1 knockdown. Development of these models implicated optimization of lentiviral transduction of suspension cell lines.
        i. Mac2A
        ii. MyLa 2000
        iii. SZ4
    b. Lung cancer cell lines with GTSF1 knockdown.
        i. H1975
    c. Overexpression model of GTSF1 for interactome analysis.
        i. HEK293T
    d. Other cell lines with GTSF1 knockdown or overexpression, not presented in here.
        i. CAL-27, knockdown.
        ii. A431, knockdown.
        iii. N/TERT-1, overexpression.
        iv. SZ4, overexpression.

Chapter III – Results

2. Transposons and their control programs are heterogeneously expressed in cell lines representing diverse malignancies.

3. *GTSF1* mRNA is ectopically expressed across diverse malignancies and its expression is mirrored in representative cell lines.
4. GTSF1 is not a master regulator of cell survival in CTCL nor lung cancer.
5. GTSF1 contributes to transposition events in lung cancer.
6. GTSF1 does not contribute to transposon-mediated genomic instability in lung cancer.
7. GTSF1 interacts with protein members of the transfer RNA pathway.

Chapter IV – Results

8. GTSF1 has a different role in each CTCL variant.
9. GTSF1 participates in T cell activation and cytokine production by CTCL cells.
10. GTSF1 mediates memory/effector phenotype of CTCL cells.
11. GTSF1 mediates phenotype changes through the NF-κB pathway.
12. GTSF1 expression in CTCL patients is heterogeneous.
13. High *GTSF1* expression in CTCL patients is associated with worse prognosis.

Contributions of author

In compliance with the Guidelines concerning thesis preparation of McGill's Graduate and Postdoctoral Studies.

Publications immediately relevant to the thesis:

1. Chapter I – Introduction, contains a summary of:

   Amelia Martínez Villarreal, Jennifer Gantchev, François Lagacé, et.al. Hypopigmented Mycosis Fungoides: Loss of Pigmentation Reflects Antitumor Immune Response in Young Patients. *Cancers (Basel).* 2020;12(8).

   I performed the literature review and wrote the manuscript.

2. Chapter III and IV – Results, contain my work from:

   Amelia Martínez Villarreal, Jennifer Gantchev, Pingxing Xie, et. al. Memory T cell phenotype in cutaneous T-cell lymphoma is modified by germline gene Gametocyte Specific Factor 1. *Manuscript under review.*

   I performed the majority of experiments, designed the research, analyzed and interpreted data, performed statistical analysis and wrote the manuscript.

Contributions of co-authors

All the figures presented in Chapters III and IV of this thesis were generated by **me**, except the following:

1. Figure 3.3A

   The Cancer Genome Atlas data retrieval and initial statistical analysis was performed by Dr. Pingxing Xie and Dr. Philippe Lefrançois. I performed the Bonferroni adjustment and created the graph.

2. Figure 3.3C

   The western blot analysis of GTSF1 expression in a panel of lung cancer cell lines was performed by Dr. Jennifer Gantchev.

3. Figure 3.3E

   The quantitative reverse transcription polymerase chain reaction (RT-qPCR) of *GTSF1* messenger RNA expression in a panel of CTCL cell lines was performed by Dr. Jennifer Gantchev. I performed expression analysis and created the graph.

Publications

In addition to the publications listed under the Contributions of Author sections, I have collaborated in the following articles.

Publications immediately relevant to the thesis

1. Fadi Touma, Marine Lambert, **Amelia Martínez Villarreal**, et. al. The Ultraviolet Irradiation of Keratinocytes Induces Ectopic Expression of LINE-1 Retrotransposon Machinery and Leads to Cellular Senescence. *Biomedicines.* 2023;11(11).

   I trained the first author to perform experiments and I reviewed and edited the manuscript.

2. Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. The Ectopic Expression of Meiosis Regulatory Genes in Cutaneous T-Cell Lymphomas (CTCL). *Frontiers in oncology.* 2019;9(429).

   I performed immunohistochemical analysis and created the figures of immunohistochemical staining.

Non-thesis related publications

3. Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. The vemurafenib paradoxical activation of MAPK pathway modifies HORMAD1 expression and genomic instability. *Manuscript under review.*

   I performed analysis from RNA sequencing differential expression results and wrote the manuscript.

4. Jennifer Gantchev, Julia Messina-Pacheco, **Amelia Martínez Villarreal**, et. al. Ectopically Expressed Meiosis-Specific Cancer Testis Antigen HORMAD1 Promotes Genomic Instability in Squamous Cell Carcinomas. *Cells.* 2023;12(12).

   I contributed to experimental planning and execution. I reviewed and edited the manuscript.

5. Philippe Lefrançois, Pingxing Xie, Scott Gun, Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. In silico analyses of the tumor microenvironment highlight tumoral inflammation, a Th2 cytokine shift and a mesenchymal stem cell-like phenotype in advanced in basal cell carcinomas. *J Cell Commun Signal*. 2020;14(2):245-254.

   I contributed to figure creation and manuscript revision.

Reviews in which I have contributed

6. Augustin Barolet, **Amelia Martínez Villarreal**, et. al. Low-Intensity Visible and Near-Infrared Light-Induced Cell Signaling Pathways in the Skin: A Comprehensive Review. *Photobiomodul Photomed Laser Surg*. 2023;41(4):147-166.

   I contributed to manuscript revision with a focus on the molecular biology.

7. Brandon Ramchatesingh, **Amelia Martínez Villarreal**, et. al. The Use of Retinoids for the Prevention and Treatment of Skin Cancers: An Updated Review. *Int J Mol Sci*. 2022;23(20).

   I contributed to manuscript revision.

8. Brandon Ramchatesingh, Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. The Contributions of Cancer-Testis and Developmental Genes to the Pathogenesis of Keratinocyte Carcinomas. *Cancers (Basel)*. 2022;14(15).

   I contributed to manuscript revision.

9. Raman Gill, Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. Understanding Cell Lines, Patient-Derived Xenograft and Genetically Engineered Mouse Models Used to Study Cutaneous T-Cell Lymphoma. *Cells.* 2022;11(4).

   I contributed to manuscript revision.

10. Jennifer Gantchev, Brandon Ramchatesingh, Melissa Berman-Rosa, Daniel Sikorski, Keerthenan Raveendra, Laetitia Amar, Hong Hao Xu, **Amelia Martínez Villarreal**, et. al. Tools used to assay genomic instability in cancers and cancer meiomitosis. *J Cell Commun Signal.* 2022;16(2):159-177.

    I contributed to manuscript revision.

11. Jennifer Gantchev, **Amelia Martínez Villarreal**, et. al. The ectopic expression of meiCT genes promotes meiomitosis and may facilitate carcinogenesis. *Cell cycle (Georgetown, Tex).* 2020;19(8):837-854.

    I contributed to literature review, figure creation and manuscript revision.

List of figures

List of tables

List of abbreviations

Abbreviation	Definition
AML	Acute Myeloid Leukemia
ANO1	Anoctamin 1
APC	Antigen presenting cell
ATM	Ataxia-telangiectasia mutated
BSA	Bovine serum albumin
CARD11	Caspase Recruitment Domain Family Member 11
CCL	C-C motif chemokine ligand
CCR	C-C chemokine receptor type
CD	Cluster of differentiation
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary DNA
CLA	Cutaneous lymphocyte-associated antigen
CT10	Cancer/Testis 10
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
CTCL	Cutaneous T-Cell Lymphoma
CXCR	C-X-C Motif Chemokine Receptor
DAB	3, 3'-diaminobenzidine
DAPI	4′,6-diamidino-2-phenylindole
DDX4	DEAD-Box Helicase 4
DEG(s)	Differentially Expressed Gene(s)
DMSO	Dimethylsulfoxide
DNMT(s)	DNA methyltransferase(s)
DSBs	Double-Strand Breaks
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme Linked Immunosorbent Assay
EOMES	Eomesodermin
FBS	Fetal bovine serum

FFPE	Formalin-fixed paraffin-embedded
GFP	Green fluorescent protein
GIN	Genomic instability
GMNN	Geminin
GTSF1	Gametocyte Specific Factor 1
HATs	Histone acetyltransferase(s)
HDACs	Histone deacetylase(s)
HDACi	HDAC inhibitors
HENMT1	HEN Methyltransferase 1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLA	Human leukocyte antigen
HMF	Hypopigmented Mycosis Fungoides
HMTases	Histone methyltransferases
HORMAD1	HORMA domain-containing protein -1
HRP	Horseradish peroxidase
H&E	Hematoxylin and eosin
IFN	Interferon
IL	Interleukin
IL1RL1	Interleukin 1 receptor-like 1
IP-Seq	Immunoprecipitation - sequencing
ITGB7	Integrin subunit beta 7
JAK	Janus kinase
LINE-1, L1	Long interspersed element - 1
lncRNA	Long non-coding RNA
MAEL	Maelstrom spermatogenic transposon silencer
MAGE1	Melanoma-associated antigen 1
MAPK	Mitogen-activated protein kinase
MF	Mycosis Fungoides
MID	Middle
miRNA(s)	MicroRNA(s)
MOI	Multiplicity of infection

MOV10L1	Mov10 Like RNA Helicase 1
mRNA	Messenger RNA
mTOR	Mammalian target of rapamycin
MTT	3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide
NF-κB, NFKB	Nuclear factor kappa-light-chain-enhancer of activated B
NK	Natural killer
NOTCH1-IC	Notch1 intracellular domain
NSG	NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ
NY-ESO-1	New York Esophageal Squamous Carcinoma-1
OE	Overexpression
ORF	Open Reading Frame
PAZ	PIWI/Argonaute/Zwille
PBS	Phosphate Buffered Saline
pcALCL	Primary cutaneous Anaplastic Large Cell Lymphoma
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDCD1	Programmed Cell Death 1
PFA	Paraformaldehyde
PI	Propidium iodide
piRNA	PIWI-interacting RNAs
PIWI	P-element-induced wimpy testis
PIWIL	PIWI like
PLCG1	Phospholipase C Gamma 1
PLD6	Phospholipase D family member 6
PMA	Phorbol 12-myristate 13-acetate
PRAME	Preferentially expressed antigen of melanoma
PUVA	Psoralen and ultraviolet A
PVDF	Polyvinylidene fluoride
p-s	Penicillin-streptomycin
RB1	Retinoblastoma 1
RIPK1	Receptor Interacting Serine/Threonine Kinase 1

RLU	Relative luminescence units
RNA-Seq	RNA Sequencing
RPA2	Replication Protein A2
RPKM(s)	Reads Per Kilobase per Million mapped reads
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
rtTA	Tetracycline-inducible transactivator
SASP	Senescence-associated secretory phenotype
SCC(s)	Squamous Cell Carcinoma(s)
SCR	Scrambled – negative control
s.d.	Standard deviation
SDS	Sodium dodecyl sulfate
shGTSF1	Short hairpin RNA GTSF1 – silenced GTSF1
shRNA	Short hairpin RNA
sRNA	Small RNA
SS	Sézary Syndrome
ssDNA	Single-stranded DNA
SSNV(s)	Somatic single nucleotide variant(s)
STAT	Signal transducer and activator of transcription
SVA	SINE-VNTR-Alu
TBST	Tris Buffered Saline with Tween 20
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TDRD9	Tudor Domain Containing 9
TEs	Transposable elements
TET	Ten-eleven translocation
Tet0	Tetracycline response promoter element
TGF-β	Transforming growth factor beta
Th	T helper type
TNF	Tumor necrosis factor
TNFR	TNF receptor 2
TNFRSF1B	TNF Receptor Superfamily Member 1B

TOX	Thymocyte selection-associated high mobility group box
TP53	Tumor protein 53
TPM(s)	Transcript Per Million
T_{RM}	Resident memory T cell
tRNA	Transfer RNA
UTR	Untranslated Region
UV	Ultraviolet
WT1	Wilms tumor 1

The following are abbreviations employed in the TCGA data in the order they appear.

LAML	Acute Myeloid Leukemia
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
LGG	Brain Lower Grade Glioma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma

PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THYM	Thymoma
THCA	Thyroid carcinoma
UCS	Uterine Carcinosarcoma
UCEC	Uterine Corpus Endometrial Carcinoma
UVM	Uveal Melanoma

Chapter 1 : Introduction

Cancer is a highly heterogeneous group of diseases characterized by uncontrolled cell growth. This means that the defining characteristics of one cancer type may not be applicable to another cancer type. As a consequence, cancer remains a challenge for health systems across the world.

Current approaches aim at focusing on the genes or proteins specific to each individual malignancy – this approach is called precision medicine. Precision medicine can take many forms, and in our lab, we have focused on the ectopic reactivation of developmental programs. In particular, we focus on reactivation of cancer-germline antigens and of transposable elements (TEs). Interestingly, the reactivation of these two developmental programs is intertwined.

1.1 Reactivation of developmental programs in cancer

Over the last few decades, researchers have been attempting to identify tumor antigens. Cancer antigens will likely trigger an antitumor immune response, making them highly relevant for the development of novel immunotherapies. Currently, tumor antigens are classified into: (1) oncovirus proteins, (2) mutated proteins, (3) fusion proteins, (4) overexpressed proteins, (5) differentiation proteins, and (6) cancer-germline antigens also called cancer/testis antigens [1,2].

One of the first identified cancer/testis antigens was termed melanoma-associated antigen (MAGE1) [3]. The gene that codes for MAGE1 was later mapped to the X chromosome. Over the next years, multiple tumor antigens were mapped to the X chromosome or were reported to be uniquely expressed in testis. Thus, the term cancer/testis antigens was coined [2]. Subsequent research demonstrated that these genes were expressed in testis as well as in fetal ovaries and in trophoblasts, therefore the term cancer-germline antigens was proposed [1,2]. Albeit some minor differences among authors, cancer-germline antigens are defined by: (1) expression in germ cells

and in a variety of tumors, (2) expression is regulated by epigenetic mechanisms, and (3) trigger of an immune response (i.e. they are immunogenic) [4,5].

Another developmental program which is commonly reactivated in cancer are TEs [6,7]. TEs are repetitive genetic elements that can move around the genome. First identified by Barbara McClintock in maize, further research identified them in multiple other organisms [6]. Under normal somatic cells, these elements are silenced and are not able to move around the genome. However, during oogenesis, spermatogenesis and early stages of development TEs are expressed and are highly abundant [6,8]. Similarly, TEs activation is now considered a biomarker of carcinogenic processes [9,10].

1.1.1 Epigenetic control of developmental programs

Expression and activation of both cancer-germline antigens and TEs is under epigenetic control. These programs are activated in germ cells and in trophoblasts, then as the organism develops and germ cells differentiate into somatic cells, these programs are silenced [11]. It is through DNA methylation and post-translational histone modifications that somatic cells control the expression of these developmental programs [11]. DNA methylation is the covalent addition of a methyl group to cytosine bases in cytosine-guanine dinucleotides of DNA (called CpG islands or CpG rich regions) by DNA methyltransferases (DNMTs) [12]. CpG islands near the promoters of some developmental genes highlight the importance of this regulating mechanism in their expression [11,13]. Germ cells and early developmental stages undergo methylating and demethylating waves depending on the developmental stage (**Figure 1.1**) [1] : Likewise, cancer cells change their DNA methylation patterns. Cancer cells undergo a genome-wide demethylation at the same time that tumor suppressor genes are silenced [9,14].

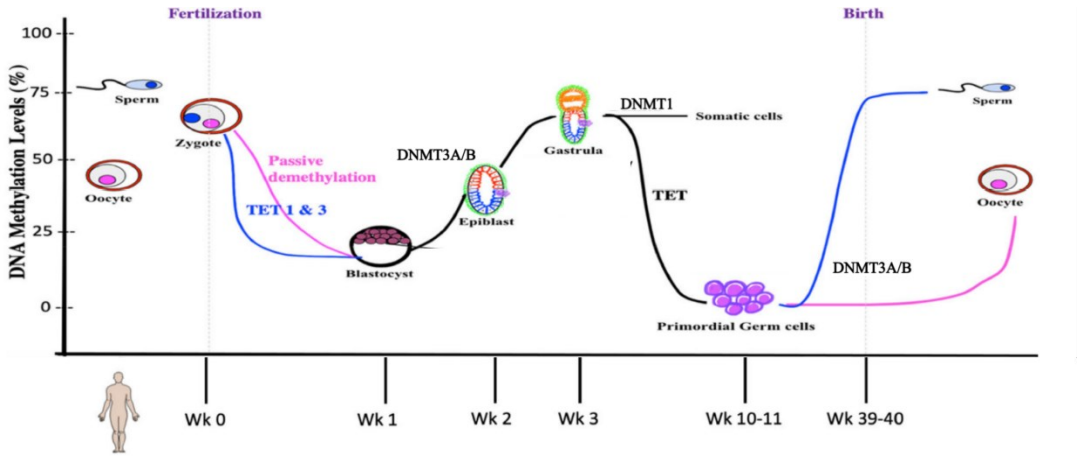


Figure 1.1 Changes in DNA methylation levels from fertilization to germ cells development.

After fertilization, TET1 and TET3 mediate passive demethylation changes in the zygote. Then at the blastocyst stage, DNMT3A and DNMT3B gradually restore methylation levels, through the epiblast until gastrula state. At the gastrula state, DNTM1 will maintain the methylation levels in somatic cells, while primordial germ cells being developed undergo another round of demethylation mediated by TET enzymes. During development of germ cells, methylation levels are restored by DNMT3A and DNMT3B. Human development timeline is presented at the bottom. DNMT, DNA methyltransferase; TET, Ten-seleven translocation; Wk, week. Adapted from Figure 1 in [15] used under CC by 4.0.

Post-translational histone modifications consist of acetylation and/or methylation of the N-terminal tails of histones forming the nucleosomes. The pattern of acetylation and methylation of histones is called histone code. Acetylation is mediated by histone acetyltransferases (HAT), that add acetyl groups, and histone deacetylases (HDAC), that removes acetyl groups [12]. Histone methylation is mediated by methyltransferases (HMTases) and demethylases [16]. Histone acetylation induces expression while methylation can do both, induce and repress expression. During their development, germ cells undergo changes in their histone code [17]. Similarly, early embryo development is associated with changes in histone acetylation patterns coordinated by expression of different HDACs (**Figure 1.2**) [18]. Treatment of cancer cells with HDAC inhibitors (HDACi) and/or demethylating agents leads to expression of developmental program genes, highlighting the role of post-translational histone modifications as an expression regulating mechanism [1,19]. The success of HDACi and demethylating agents for treatment of cancer evidence the role epigenetics play in carcinogenesis [20].

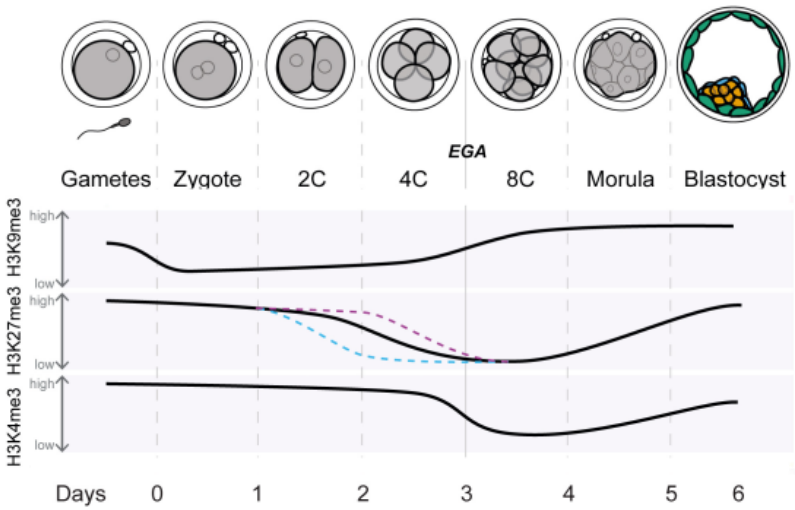


Figure 1.2 Changes in levels of different histone marks from gamete to blastocyst.

The levels of histone marks H3K9me3, H3K27me3, and H3K4me3 change during embryogenesis. The black lines indicate changes in global levels, while the pink and blue dashed lines indicate changes in maternal and paternal genomes respectively. 2C, 2-cell stage; 4C, 4-cell stage; 8C, 8-cell stage; EGA, embryonic genome activation. Adapted from Figure 1 in [21] used under CC by 4.0.

1.1.2 Reactivation of cancer-germline antigens in cancer

The identification of cancer-germline antigens led to the hypothesis that these genes contribute to carcinogenesis. Currently, there is a plethora of evidence to presume cancer-germline antigens actively contribute to every hallmark of cancer (**Figure 1.3**). Cancer-germline antigens can regulate transcriptional programs that lead to carcinogenesis, cell division, genomic instability (GIN), DNA damage response, apoptosis, invasion and metastasis, cell energetics, autophagy and modification of the microenvironment [11,13]. For example, MAGE1 counteracts Notch1 intracellular domain (NOTCH1-IC) by recruiting HDAC1, thus leading to repression of differentiation programs and contributing to carcinogenesis [2]. Recently in our lab we demonstrated the role of HORMA domain-containing protein -1 (HORMAD1) in maintaining a level of GIN that is beneficial for malignant cells in squamous cell carcinomas (SCCs) [22].

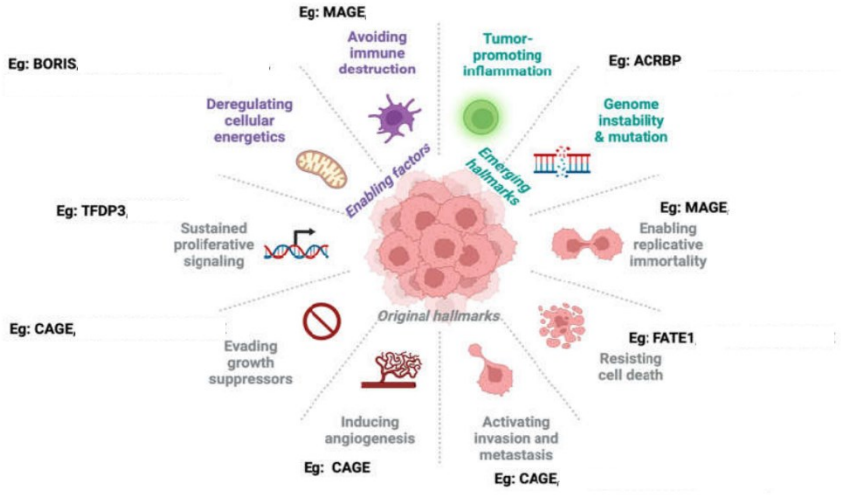


Figure 1.3 Evidence of the role cancer-germline antigens have in the hallmarks of cancer.
Hallmarks of cancer and examples of cancer-germline antigens that participate in each hallmark. In the center of the circle hallmarks are listed and classified according to [23], with enabling factors in purple, emerging hallmarks in aqua and the original hallmarks in grey. In the edges of each slice, examples of cancer-germline antigens that contribute to each hallmark are provided. ACRBP, Acrosin Binding Protein; BORIS, Brother Of The Regulator Of Imprinted Sites; CAGE, Cancer-Associated Antigen; FATE1, Fetal And Adult Testis Expressed 1; MAGE, Melanoma-Associated Antigen; TFDP3, Transcription Factor Dp Family Member 3. Adapted from Figure 2 in [11] used under CC by 4.0.

In addition, it has been suggested that the reactivation of cancer-germline antigens in cancer allows the malignant cells to acquire stem cell properties. Acquisition of stem cell properties enables tumor maintenance, proliferation and metastasis [4]. Such hypothesis suggests that cancer-germline antigens would be damaging for normal somatic cells, but cancer cells are able to take advantage of their expression [4]. Expression of cancer-germline antigens then, would be orchestrated by genes that control germ cell gene expression, such as those controlling epigenetic modifications [2]. Furthermore, it is speculated that the contradictory signals in cancer cells to express and to repress germ cell genes, would lead to a heterogeneous expression of cancer-germline antigens in tumors [1].

1.1.3 Reactivation of transposon elements in cancer

TEs constitute up to 50% of the human genome and considering they are able to move around the genome, these elements constitute a latent danger [14]. TEs are classified based on their mechanism of movement, called transposition. Class I moves through a mechanism termed copy-and-paste in which they use reverse transcribed RNA as an intermediate to insert into the genome (**Figure 1.4**). Due to the reverse transcription step, these TEs are often called retrotransposons. Class II elements move through a mechanism termed cut-and-paste in which they excise from the donor sequence and reintegrate into another area of the genome [10] (**Figure 1.4**). The only active transposon in humans is Long interspersed element - 1 (LINE-1 or L1). L1 consists of two open reading frames (ORF), ORF1 and ORF2, that encode for a nucleic acid chaperone [24] and an endonuclease and reverse transcriptase, respectively [9] (**Figure 1.5**). L1 is able to transpose itself (*cis* transposition) and other TEs (*trans* transposition), such as Alu elements and SINE-VNTR-Alu (SVA) elements [9].

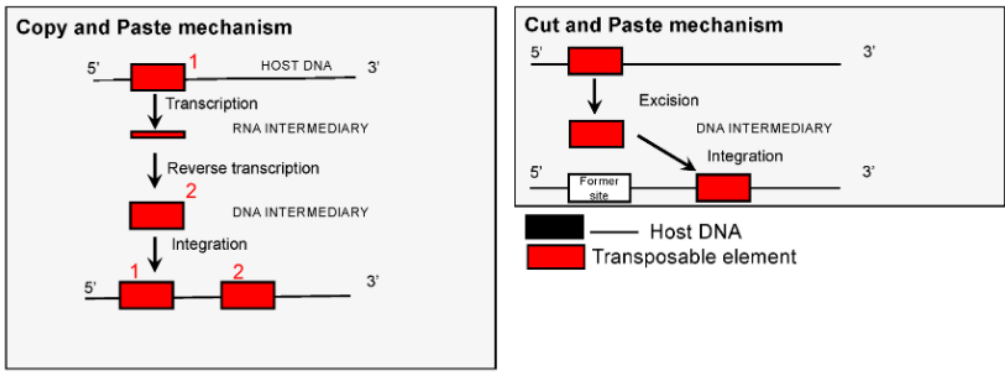


Figure 1.4 Mechanisms of transposition: copy-and-paste and cut-and-paste.

In the copy-and-paste mechanism shown on the left box, TEs produce an RNA intermediary by transcription, followed by reverse transcription to insert a new copy of the complementary DNA in the host genome. Due to the reverse transcription step, these TEs are often called retrotransposons. Numbers in red indicate the original (1) and new (2) insertion. In the cut-and-paste mechanism shown on the right box, the transposon is excised from the host genome and is inserted in a new position in the genome. Adapted from Figure 1A in [25] used under CC by 4.0.

Transposon reactivation in cancer has been studied at the mRNA, protein and functional level [9,10]. However, how new transposon insertions (i.e. when they become functional) contribute to carcinogenesis is better understood. New transposon reinsertions can cause transcriptional deregulation, GIN, chromosomal rearrangements, inactivation of tumor suppressor genes, activation of oncogenes and alterations in non-coding RNAs [14]. For example, a retrotransposon insertion in the gene Adenomatous polyposis coli disrupted its ORF contributing to the development of colorectal cancer [26]. New transposon insertions can cause GIN by insertional mutagenesis, induction of unstable microsatellites, alterations in transcription rate, disruption of DNA repair systems and by chromosomal rearrangements [14].

Figure 1.5 L1 structure.

L1 is constituted of two ORF. ORF1 is formed by a coiled-coil domain, an RNA recognition motif and a carboxyl-terminal domain. ORF2 is formed by an endonuclease domain, a reverse transcriptase domain and a cysteine-rich domain. C, cysteine-rich domain; CC-LZ, coiled-coil domain; CTD, carboxyl-terminal domain; EN, endonuclease domain; ORF, Open Reading Frame; RRM, RNA recognition motif; RT, reverse transcriptase domain. Adapted from Figure 4 in [27] used under CC by 4.0.

1.1.4 Advantages of targeting developmental programs in cancer

The advantages of targeting reactivated developmental programs in cancer derive primarily from the potential of these programs to produce antigens. These antigens can then be presented by Human Leukocyte Antigen (HLA) Class I molecules in the surface of the tumor cell or by HLA Class II on the surface of antigen presenting cells (APC); then, an immune response against the antigen and, consequently, the tumor is activated [19].

The identification of the tumor antigens as something foreign is aided by the immune privilege testis, fetal ovaries and trophoblasts have. Immune privilege refers to the inability of the immune system to develop tolerance to self-antigens in that tissue. Immune privilege can be obtained by multiple mechanisms, such as the blood-tissue barrier, lack of lymphatic circulation and immune cell-mediated mechanisms [28]. This means that, in theory, when a cancer cell or an APC presents an antigen derived from a reactivated developmental program to a T cell, the T cell would be able to recognize it as foreign and activate an immune response against it.

Some of these antigens have proved immunogenicity in patients. For example, antibodies against Cancer/Testis 10 (CT10) in a melanoma patient were identified [29], as well as humoral and cell-mediated immune responses against New York Esophageal Squamous Carcinoma-1 (NY-ESO-1) in lung, ovarian, breast, melanoma and esophageal cancer patients [12]. In fact, there is a correlation between the expression of these genes and the presence of B and T cells specific for antigens derived from these genes [13].

Both, cancer-germline antigens and TEs produce tumor antigens, therefore therapeutic approaches targeting both have been developed. The two main approaches to target reactivation of developmental programs are cancer vaccination and adoptive transfer of antigen-specific T cells

(**Figure 1.6**). [30,31]. The objective of both approaches is to elicit an immune response that leads to elimination of cancer cells. Vaccination stimulates the antitumor immune response of the patient thereby inducing an active immunity; adoptive transfer infuses the patient with T cells that are specific for the antigen, inducing a passive immunity [31]. In addition, it has been suggested to employ epigenetic therapies to induce TEs expression that can be targeted by any of these immunotherapies [30].

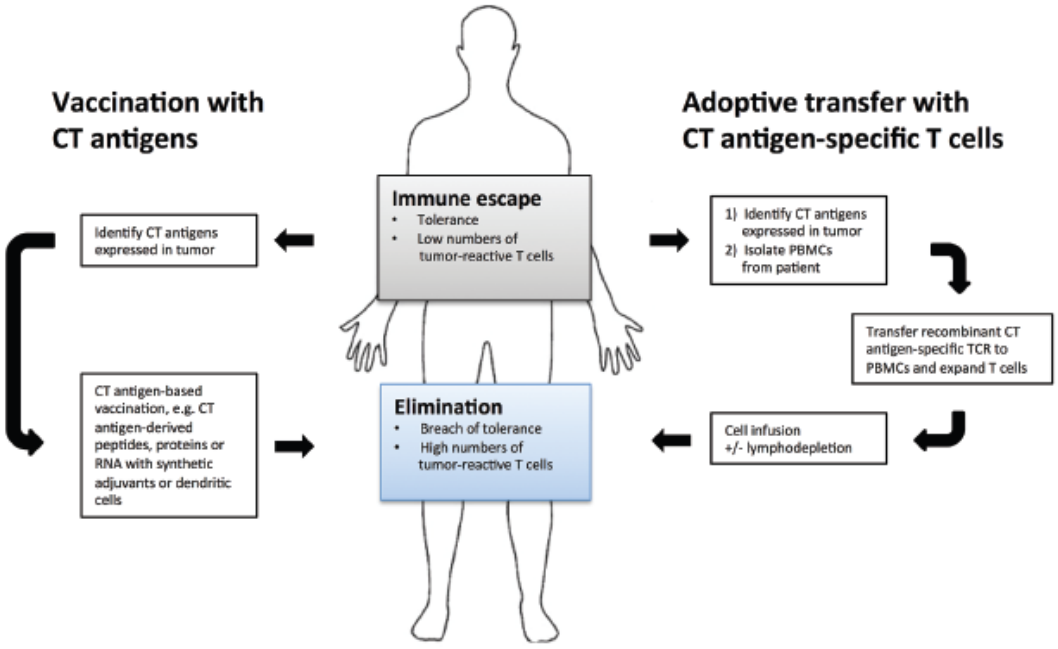


Figure 1.6 Immunotherapy targeting cancer-germline antigens and TEs.

Two main approaches to target reactivation of developmental programs: vaccination and adoptive transfer of antigen-specific T cells. Both approaches try to eliminate cancer cells that have been able to escape the immune system recognition. Vaccination induces a response from the host system while adoptive transfer infuses the cells that will respond to the cancer. The steps needed to develop each immunotherapy are detailed. Overall, identification of a feasible antigen is key. CT, cancer testis; PBMCs, peripheral blood mononuclear cells; TCR, T Cell Receptor; TEs, transposable elements. Copied from Figure 3 in [31] used under CC by 4.0.

1.2 The piRNA pathway

The reactivation of the two developmental programs introduced above, TEs and cancer-germline antigens, is intertwined. During germ cell development, epigenetic changes lead to TEs

expression. Therefore, germ cells developed mechanisms to counterbalance TEs expression. Interestingly, members of these mechanisms are often ectopically expressed in cancer and, thus, classified as cancer-germline antigens.

One of the mechanisms developed by cells for TEs control is the P-element-induced wimpy testis (PIWI)-interacting RNA (piRNA) pathway. In germ cells, the piRNA pathway controls TEs reactivation during the developmental epigenetic reprogramming [8,32].

1.2.1 Elements of the piRNA pathway

The members of the piRNA pathway are the PIWI protein family and the piRNAs. Each PIWI protein interacts with a specific population of piRNAs. The PIWI protein family in mouse is constituted by MILI, MIWI and MIWI2; PIWI like 2 (PIWIL2), PIWIL1 and PIWIL4 are the human orthologs, respectively. All PIWI proteins are members of the Argonaute family and present PIWI, middle (MID) and PIWI/Argonaute/Zwille (PAZ) domains. The PIWI domain has endonucleolytic slicer activity, the MID domain anchors the 5' and the PAZ domain the 3' of piRNAs [8]. PiRNAs are a type of small RNAs (sRNAs) and can be TEs, messenger RNA (Mrna) or long non-coding RNA (lncRNA)-derived; TE-derived piRNAs are transcribed from both genomic strands by Pol II generating 24-32 nucleotide long RNAs [33,34].

1.2.2 PiRNA biogenesis

In mice, after transcription, piRNAs are exported to the cytoplasm by Maelstrom spermatogenic transposon silencer (MAEL). Once in the cytoplasm, piRNAs bind to the RNA helicase Mov10 Like RNA Helicase 1 (MOV10L1). MOV10L1 facilitates the interaction with Phospholipase D family member 6 (PLD6), which trims the 5' end of the piRNA [8]. The 5' processed end allows interaction with MILI or MIWI2 for stabilization of the piRNA. Then an

unidentified trimmer processes the 3' end of the piRNA. Finally, mature 3' ends are methylated by HEN Methyltransferase 1 (HENMT1), to provide stability and protection to the piRNA [32]. This process is called primary piRNA biogenesis [8] (**Figure 1.7**).

In secondary piRNA biogenesis, primary piRNAs are used as templates to produce secondary piRNAs derived from their antisense transcripts. MILI binds secondary piRNAs by complementarity to their already bound primary piRNA. This interaction allows processing and trimming of the secondary piRNA. Then, secondary piRNAs fuel the processing and trimming of primary piRNAs, therefore this step has been termed the ping-pong cycle (**Figure 1.7**) [8].

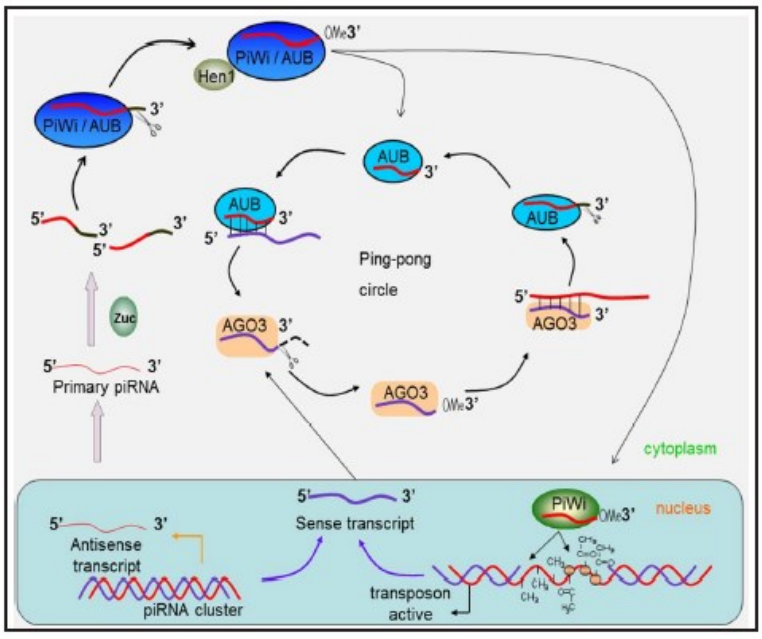


Figure 1.7 piRNA biogenesis.

Simplified diagram of primary and secondary (ping-pong) piRNA biogenesis in *Drosophila*. The process in mice follows the same principle. Zuc is the *Drosophila melanogaster* ortholog of mouse PLD6, AUB is the *Drosophila melanogaster* ortholog of mouse MIWI and AGO3 is a *Drosophila melanogaster* analog to mouse PIWI proteins. AGO3, Argonaute 3; AUB, Aubergine; piRNA, PIWI-interacting RNAs; PIWI, P-element-induced wimpy testis; Zuc, zucchini. Copied from Figure 1 in [33] used under CC by nc nd 4.0.

Finally, MIWI2 bound to a piRNA translocate to the nucleus where by complementarity will identify active TEs in the genome. This identification recruits DNMTs and HMTases for DNA methylation and H3K9me3 thereby silencing the active TEs (**Figure 1.8**) [8].

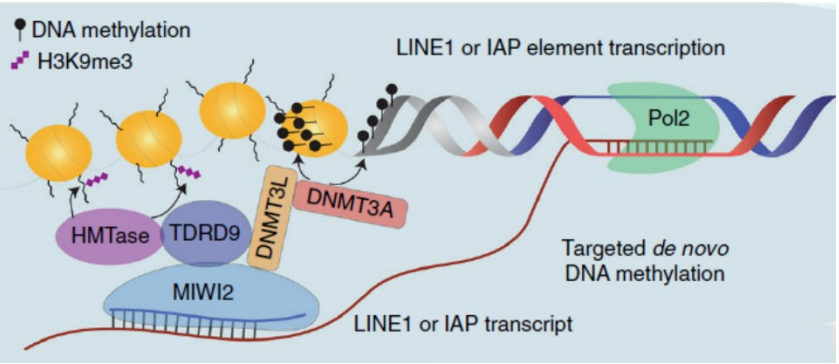


Figure 1.8 Epigenetic silencing of active TEs by the piRNA pathway.

Once translocated to the nucleus, MIWI2 and TDRD9 complex loaded with a piRNA, identifies transposon RNA produced by Pol II. Then, MIWI2 recruits DNMTs and HMTases, leading to DNA methylation and histone modification H3K9me3. Consequently, active TEs are now in heterochromatin state, inhibiting their expression. DNMT3A, DNA methyltransferase 3 A; DNMT3L, DNA methyltransferase 3 like; HMTase, Histone methyltransferases; IAP, Intracisternal A particle; LINE1, Long interspersed element - 1; piRNA, PIWI-interacting RNAs; Pol2, polymerase 2; TDRD9, Tudor Domain Containing 9; TEs, transposable elements. Adapted from Figure 2 in [8] used under CC by 4.0.

1.2.3 Expression of piRNA elements in cancer

Despite the piRNA pathway being a germ cell-specific mechanism, expression of both PIWI proteins and piRNAs in cancer has been reported. They have been reported to be involved in proliferation, apoptosis, invasion and metastasis. Interestingly, the interaction between PIWI proteins and piRNAs in a cancer background has been less studied [33]. A few examples include the downregulation of piR-55490 in lung cancer which leads to increased activation of mammalian target of rapamycin (mTOR) and the upregulation of PIWIL2 in cervical SCC where it inhibits apoptosis [33].

1.3 Gametocyte Specific Factor 1

Gametocyte Specific Factor 1 (GTSF1) is a piRNA pathway member that has been reported to be expressed in cancer. It is a 167-amino acid protein that is essential for spermatogenesis. Interestingly, GTSF1 is a small protein (~19 kDa) with a simple structure, two N-terminal CHHC zinc fingers. In addition, consistent with other piRNA pathway elements, GTSF1 is highly conserved from insects to humans [35]. Orthologs of GTSF1 have been extensively studied in *Drosophila* (Asterix/DmGTSF1) and in the silkworm *Bombyx mori* (BmGtsf1), but additional orthologs have been identified in metazoans and protozans (**Figure 1.9**).

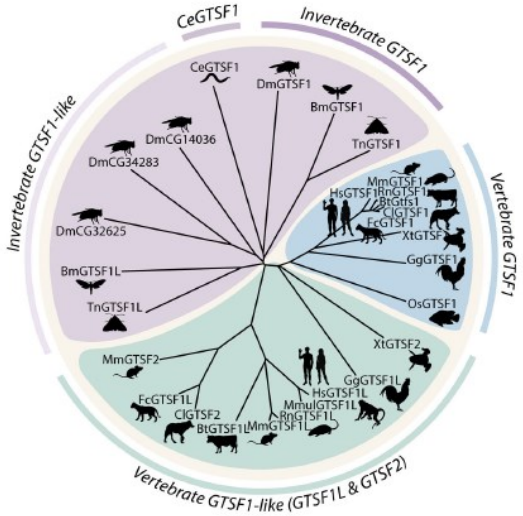


Figure 1.9 Phylogenetic analysis of GTSF1 orthologs and GTSF1-like orthologs.
In light blue are shown GTSF1 orthologs identified in vertebrates, while in green other vertebrates show expression of GTSF1-like orthologs. Invertebrates in purple show greater differences. Bm, *Bombyx mori*; Bt, *Bos taurus*; Ce, *Caenorhabditis elegans*; Cl, *Canis lupus familiaris*; Dm, *Drosophila melanogaster*; Fc, *Felis catus*; Gg, *Gallus gallus*; GTSF1, Gametocyte Specific Factor 1; GTSF1L, Gametocyte Specific Factor 1 Like, GTSF2, Gametocyte Specific Factor 2; Hs, *Homo sapiens*; Mmul, *Macaca mulatta*; Mm, *Mus musculus*; On, *Oreochromis niloticus*; Rn, *Rattus norvegicus*; Tn, *Trichoplusia ni*; Xt, *Xenopus tropicalis*. Adapted from Figure 6C in [35] used under CC by 4.0.

GTSF1 was initially identified when Yoshimura and colleagues were searching for a transcript that was exclusively expressed in unfertilized eggs, ovaries, and testes of mice [36]. After

several years of this initial report, GTSF1 expression in human testis, fetal ovary, oocytes, and preimplantation embryos was reported [37,38].

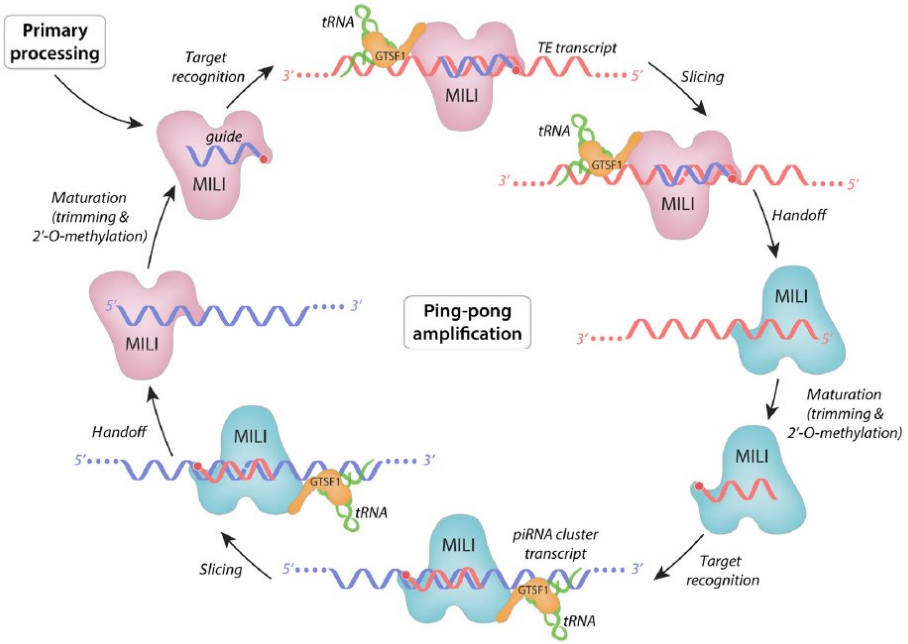


Figure 1.10 GTSF1 participation in the secondary piRNA biogenesis.
The current model of GTSF1's participation in the secondary piRNA biogenesis also known as the ping-pong amplification. GTSF1 participates by grasping and stabilizing piRNAs bound to PIWI proteins. The interaction is aided by tRNA bound to GTSF1. GTSF1, Gametocyte specific factor 1; piRNA, PIWI-interacting RNA; TE, transposable element; tRNA, transfer RNA. Copied from Figure 2 in [39] used under CC by 4.0.

1.3.1 The role of GTSF1 in germ cells

GTSF1 knockout male mice appear normal but are sterile due to apoptotic death of their germ cells [40]. This phenotype was similar to other piRNA pathway mutants suggesting GTSF1 belonged to this pathway. Further analysis concluded that GTSF1 participated in the secondary piRNA biogenesis [39]. Specifically, GTSF1 grasps and stabilizes piRNAs so they can be processed by PIWI proteins (**Figure 1.10**). Furthermore, GTSF1 binds to transfer RNAs (tRNAs) to help identify active TEs in the genome [41]. Specifically, it was proposed that GTSF1, as a member of

the RNA silencing complex of the piRNA pathway, aids in the recruitment of the silencing machinery to the nucleus (**Figure 1.11**).

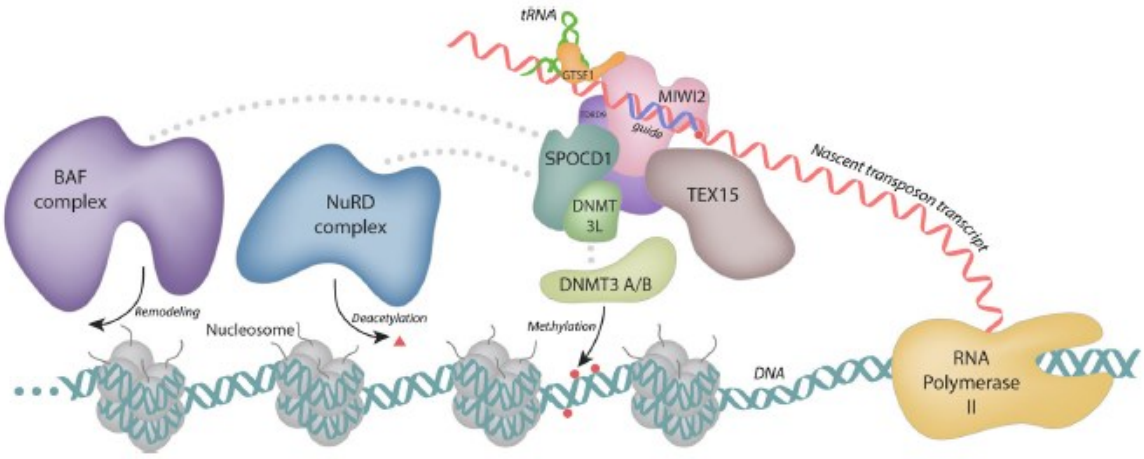


Figure 1.11 GTSF1 participation in recruitment of silencing machinery.
Current model of GTSF1 participation in post-transcriptional silencing of TEs as part of the piRNA pathway. Mature silencing complexes, MIWI2 bound to piRNA in concert with GTSF1 associated with tRNA, identify active TEs. This triggers the recruitment of silencing machinery, DNMTs and histone remodelers. BAF, Brahma-associated factor; DNMT, DNA methyltransferase; GTSF1, Gametocyte Specific Factor 1; nuRD, Nucleosome Remodeling and Deacetylase; SPOCD1, SPOC Domain Containing 1; TDRD9, Tudor Domain Containing 9; TEs, transposable elements; TEX15, Testis expressed 15; tRNA, transfer RNA. Adapted from Figure 4 in [35] used under CC by 4.0.

1.3.2 The role of GTSF1 in cancer

Ectopic GTSF1 expression has been reported in Cutaneous T-Cell Lymphoma (CTCL) [42-49], Acute Myeloid Leukemia (AML) [50,51] and liver cancer [52].

The first report of *GTSF1* expression in CTCL analyzed publicly available datasets of gene expression and compared the CTCL variant Mycosis Fungoides (MF) tumor stage to normal skin, inflamed skin and normal T cells. The authors reported *GTSF1* is significantly upregulated in MF tumor stage compared to their controls [47]. Then, a series of publications from our lab analyzed GTSF1 expression, along with other potential biomarkers, with different methodologies: (1) with quantitative reverse transcription polymerase chain reaction (RT-qPCR) patient samples were

analyzed and confirmed *GTSF1* expression; furthermore, patient-derived cell lines demonstrated GTSF1 protein expression could be modified with HDACi [42]; (2) with RT-qPCR from patient samples ectopic expression of multiple genes was associated with worse prognosis, *GTSF1* was considered among this gene cluster [43]; (3) with targeted RNA-Sequencing (RNA-Seq) ectopic expression of multiple genes, high *GTSF1* among them, in advanced disease stages was validated [44]; and (4) targeted RNA-Seq was repeated with samples taken more recently demonstrating stage IV CTCL had a *GTSF1* expression Log_2 ratio of 11.19893 compared to benign dermatoses [49]. More recently, with single cell RNA-Seq *GTSF1* expression was identified in clonally expanded T cells from patient samples, further confirming expression by malignant cells (**Figure 1.12**) [46]. Additionally, the molecular profiles for CTCL patients proposed, T central memory and cytotoxic effector memory T cell phenotype profiles, were associated with *GTSF1* expression; specifically, the T central memory profile [45]. Furthermore, isolation of cancer cells with laser capture microdissection followed by transcriptomic analysis demonstrated upregulation of *GTSF1* in stages ≥ IIB [48]. Therefore, ectopic *GTSF1* expression in CTCL and its association with progression and a worse prognosis has been recurrently reported.

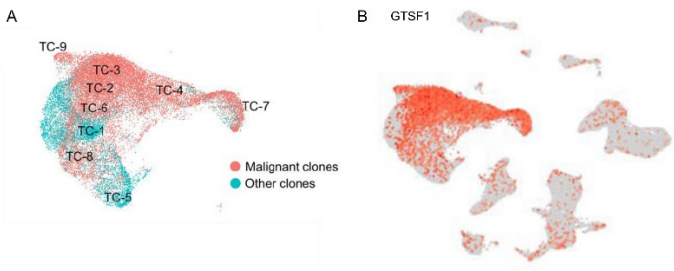


Figure 1.12 GTSF1 in CTCL.

A. Single cell RNA-Seq analysis of T cell clusters from skin patient samples. Cells are colored according to the clonality of their TCR, with malignant cells in red and other polyclonal cells in aqua. **B.** Feature plot showing that the majority of malignant clones express GTSF1, denoted in red. GTSF1, Gametocyte specific factor 1; TC, T cells; TCR, T cell receptor. Adapted from Figure 2C and Figure 2N in [46] used under CC by 4.0.

In the case of AML, gene expression patterns of patient samples with Wilms tumor 1 (*WT1)* mutation were analyzed and showed the most upregulated gene was *GTSF1* [50]. In addition, an analysis to identify microRNAs (miRNAs) associated with AML survival identified hsa-miR-589. Further analysis identified *GTSF1* as a mRNA target of hsa-miR-589. In the case of liver cancer, *GTSF1* expression in hepatocellular carcinoma tissue was higher than their normal adjacent tissue. Silencing of GTSF1 in an hepatoma cell line led to decreased cell proliferation and formation of smaller tumors [52].

Interestingly, despite the multiple reports of ectopic *GTSF1* expression in cancer tissues, no functional analysis has been performed. The recurrent report of *GTSF1* expression in CTCL by our lab and others suggests that, at least in this cancer type, GTSF1 is playing an important role in carcinogenesis.

1.4 Cutaneous T-Cell Lymphoma

Skin cancer is often classified in two broad categories: nonmelanoma skin cancer and cutaneous malignant melanomas. These cancer types arise from different cell types in the skin. Nonmelanoma skin cancer can be further subdivided into basal cell carcinoma, which arises from basal cells in the epidermis and into SCC, which arises from keratinocytes in the upper layers of the epidermis. Cutaneous melanomas arise from melanocytes [53]. Often classified as a rare cancer, CTCL is a cancer that arises from T cells but presents clinically in the skin [54].

CTCL is a rare extranodal non-Hodgkin lymphoma characterized by the expansion of malignant T cells within the skin [54]. In 2018, the annual age-adjusted incidence of CTCL in the United States was of 6.4-9.6 cases per million people [55]. CTCL more commonly presents in older male individuals [56,57]. Although the etiopathogenesis of CTCL remains an open question, most

researchers agree that this malignancy arises in a background of chronic antigen stimulation [58]. Some proposed etiological entities include medications, viruses, and *Staphylococcus aureus* infection [59].

1.4.1 Definition of CTCL variants

CTCL is a highly heterogeneous malignancy and as such, it can be further subdivided into multiple variants. The most common variants are MF, Sézary Syndrome (SS) and primary cutaneous Anaplastic Large Cell Lymphoma (pcALCL). These three variants constitute approximately 80% of all CTCL cases [60]. Each variant has its own set of clinical and/or molecular characteristics which are further described below (**Figure 1.13**).

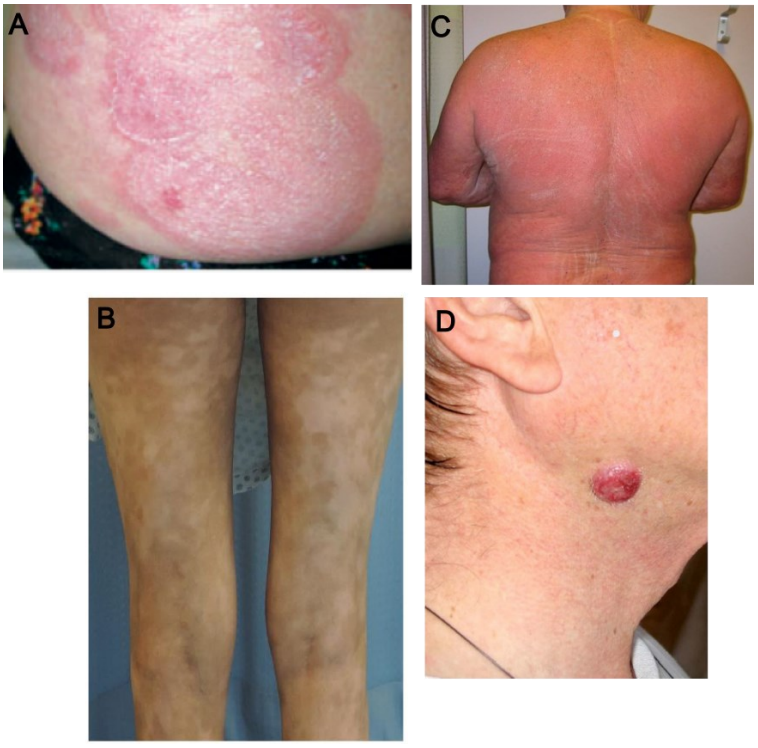


Figure 1.13 Different clinical presentations of CTCL.

A. Patch and plaques Classic MF affecting the lower trunk, **B.** Hypopigmented MF affecting the back of the legs, **C.** SS in the back showing generalized erythroderma, **D.** pcALCL presenting as a single erythematous nodule in the neck. Panel A was adapted with permission Figure 1A from [61] copyright Massachusetts Medical Society. Panel B was adapted

from Figure 1 in [62] used under CC by 4.0. Panel C was adapted from Figure 2 in [63] used under CC by nc 3.0. Panel D was adapted from Figure 11A in [64] used under CC by 4.0.

MF can be further subdivided into different variants, such as Classic MF, hypopigmented MF, folliculotropic MF, pagetoid reticulosis, among others [65]. Classic MF is defined by the presence of erythematous patches, plaques or tumors in non-sun exposed areas. These lesions often present pruritus and are scaly [56,65,66]. SS is characterized by a more aggressive course, erythroderma, pruritus and blood involvement. The current criteria to identify SS is the presence of at least 1,000 Sézary cells per cubic millimeter of blood. Sézary cells are circulating malignant T cells with an enlarged cerebriform nuclei [61,65,67]. SS can arise *de novo* or can progress from MF; in fact, it is still debated whether MF and SS are two extremes of the same disease or are completely different clinical entities [65,66]. PcALCL patients usually present a single lesion or a cluster of small tumors and malignant cells express the cell surface receptor Cluster of Differentiation (CD) 30 [68].

1.4.2 Diagnosis

Diagnosis of CTCL is challenging, particularly in early disease stages, due to resemblance to benign dermatoses [56]. Consequently, diagnosis can take up to 3-4 years [69]. The diagnosis of CTCL relies heavily on clinicopathological characteristics of the patient. In addition to a complete body assessment to estimate the percentage of body surface area involved, diagnostic and staging tests include blood cell counts, lactate dehydrogenase and skin biopsy [70]. Histopathology analysis of skin biopsies is key for establishing an accurate diagnosis. In addition, skin biopsies are often subjected to clonality assessments through polymerase chain reaction (PCR) or high-throughput sequencing. Determination of clonality has been proposed as a diagnostic test to differentiate from benign dermatoses [67].

1.4.3 Histopathology

Histopathological features are essential for diagnosing this malignancy (**Figure 1.14**). CTCL is characterized by skin infiltration of malignant T cells. Specifically, T cells migrate to the dermis and the epidermis, often accumulating around Langerhans cells to form what are called Pautrier's microabscesses. Tumor lesions often present less epidermotropism but more dermal infiltration [65,70]. In contrast, epidermotropism in pcALCL is subtle and cells accumulate around blood vessels [64]. Malignant T cells are pleomorphic enlarged cells with an irregular nucleus [56,61,65].

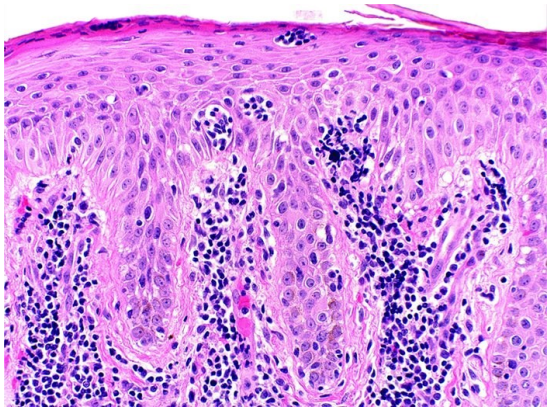


Figure 1.14 Histopathology of Mycosis Fungoides, hematoxylin and eosin staining.
Histopathology of CTCL variant, Mycosis Fungoides, showing high lymphocyte infiltration (intense purple cells) both in the epidermis and dermis. Note the lymphocyte infiltration at the upper layer of the epidermis. Copied from Figure 2 in [71] used under CC by 4.0.

In addition to routine hematoxylin and eosin (H&E) staining, skin biopsies are often stained for other markers. Malignant cells are often $CD4^+$ and cases of $CD8^+$ expression are rare. Loss of CD2, CD5 and CD7 expression is often reported [56,61,65]. In pcALCL, CD30 expression is present in >75% of malignant T cells (**Figure 1.15**) [64].

1.4.4 Treatment

Currently, there is no curative treatment for CTCL. Therefore, the goal of treatment is the control of symptoms and postponing progression while maintaining quality of life. Treatment

depends largely on the stage of the disease. Skin-directed therapies are preferred for early disease stages, while systemic therapies are preferred for late disease stages [54,70].

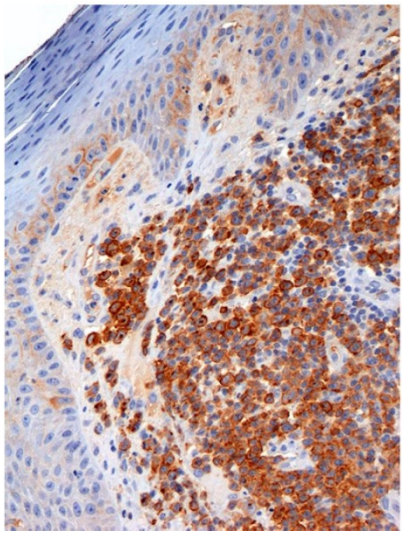


Figure 1.15 Histopathology of pcALCL, immunohistochemistry for CD30.
Malignant T cells in pcALCL are $CD30^+$. CD, Cluster of differentiation; pcALCL, Primary cutaneous Anaplastic Large Cell Lymphoma. Adapted from Figure 15A in [64] used under CC by 4.0.

Early disease stages have a favorable prognosis with complete response rates ranging between 60% to 100% [65]. Topical corticosteroids, topical retinoids (e.g. bexarotene and tazarotene) and topical nitrogen mustard are skin-directed therapies that induce apoptosis, DNA damage and affect cell proliferation and differentiation [54,65,70]. Another common treatment in early disease stages is photochemotherapy with psoralen and ultraviolet (UV) A (PUVA). PUVA consists of oral intake of 8-methoxypsoralen followed by exposure to UV A light. UV A light exposure activates the psoralen leading to DNA cross-linking and reactive oxygen species formation [72]. Excision is often the treatment of preference for pcALCL [68].

Advanced disease stages can be treated with chemotherapy, biologic or targeted therapies, but relapses are frequent. The response rates vary, depending on the treatment option selected [65]. Methotrexate and pralatrexate are antifolate chemotherapies that inhibit the metabolism of

malignant cells leading to apoptosis. Interferon (IFN) α is a biologic treatment that induces a T helper type (Th) 1 immune response in patients. Extracorporeal photopheresis follows the same principle than PUVA. Circulating mononuclear cells are extracted from the patient and isolated with a leukapheresis-based method. Then cells are mixed with 8-methoxypsoralen, exposed to UVA light, and reinfused back to the patient. Some targeted therapies that have been approved for CTCL are alemtuzumab, a humanized monoclonal antibody against CD52; HDACi, such as vorinostat and romidepsin, that cause changes in gene expression; denileukin diftitox, a fusion protein of interleukin (IL) -2 with diphtheria toxin; brentuximab vedotin, an antibody-drug conjugate that targets CD30; and mogamulizumab, an anti-C-C chemokine receptor type (CCR) 4 monoclonal antibody [54,65,70].

Hematopoietic stem cell transplantation is the only treatment with curative potential, but relapses are frequent and treatment mortality ranges between 25% to 30%. Thus, this treatment option is usually considered only for young patients [65,70].

1.4.5 Molecular features

The molecular features of CTCL have been slowly identified over the years. Although a molecular classification, as in other cancer types, has not been attained, there is some key molecular understanding about the behavior of this malignancy. Epidermotropism is one of the defining characteristics of CTCL. Malignant cells are able to migrate and home in the skin thanks to expression of cell surface molecules that interact with their corresponding ligands expressed by keratinocytes and other bystander cells (**Figure 1.16**). CCR4 expressed by malignant cells allows interaction with basal keratinocytes and endothelial cells in the skin that express C-C motif chemokine ligand (CCL) 17; CLA expressed by malignant cells allows interaction with endothelial cells in the skin that express E-selectin [61].

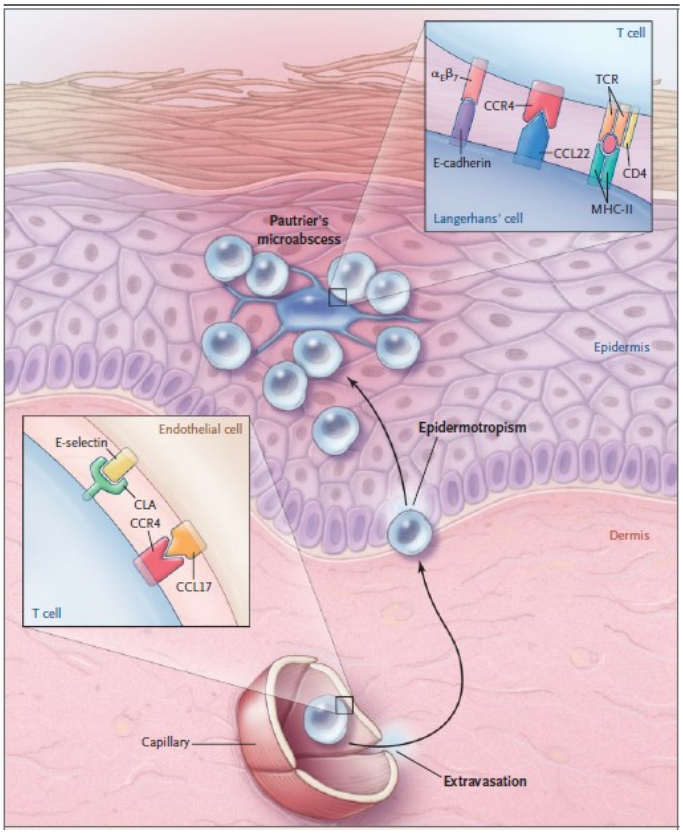


Figure 1.16 Molecular features that allow epidermotropism in CTCL.

Malignant cells are able to circulate in capillaries thanks to the expression of CLA and CCR4 that bind to E-selectin and CCL17, respectively, expressed by endothelial cells. In addition, malignant cells are attracted towards Langerhans cells due to the expression of E-cadherin, CCL22 and MHC-II by Langerhans cells. The corresponding receptors in malignant T cells are integrin $\alpha_5\beta_7$, CCR4 and TCR. These interactions enable the defining characteristic of epidermotropism. CCR, C-C chemokine receptor type; CCL, C-C motif chemokine ligand; CD, Cluster of differentiation; CLA, Cutaneous lymphocyte-associated antigen; MHC-II, Major histocompatibility complex Class II. Reproduced with permission Figure 2 from [61], copyright Massachusetts Medical Society.

Analysis of the somatic single nucleotide variants (SSNVs) in CTCL revealed this malignancy present a mutational signature of UV light exposure. This mutational signature means that approximately 75% of SSNVs are C>T transitions, which are commonly caused by UV light exposure. Considering that the skin is frequently exposed to UV, this suggests that malignant cells were already homing to the skin before acquiring the mutational signature [66,73].

Mutations that contribute to CTCL carcinogenesis can be classified according to the gene types: (1) mutations in canonical cancer genes, particularly those associated with DNA damage repair, cell cycle, apoptosis, Mitogen-activated protein kinase (MAPK) pathway and chromatin modifying genes, and (2) mutations in T cell lineage specific genes, particularly those in the T Cell Receptor (TCR), Nuclear factor kappa-light-chain-enhancer of activated B (NF-κB) and Janus kinase (JAK)- Signal transducer and activator of transcription (STAT) signaling pathways [66]. Commonly mutated canonical cancer genes include cyclin-dependent kinase inhibitor 2A (*CDKN2A*), retinoblastoma 1 (*RB1*), Ataxia-telangiectasia mutated (*ATM*) and Tumor protein 53 (*TP53*) [73]. In addition, TCR mutations are common in CTCL. These mutations might constitutively activate TCR signaling or disable the negative regulations that controls TCR signaling leading to differentiation, growth and survival [73]. For example, there are recurrent gain of function mutations in *CD28*, a TCR costimulatory molecule. These gain of function mutations lead to increased affinity of CD28 for CD86 leading to hyperactivation of the TCR signaling pathway. Other common mutations in TCR signaling genes in CTCL include Cytotoxic T-Lymphocyte Associated Protein 4 (*CTLA4)-CD28* gene fusions, Phospholipase C Gamma 1 (*PLCG1*) gain of function mutations and Programmed Cell Death 1 (*PDCD1*), the coding gene for PD-1 receptor, deletions [66,73].

The NF-κB signaling pathway is often constitutively activated in CTCL. Activation of this pathway leads to production of proinflammatory cytokines, proliferation, inhibition of apoptosis and activation of differentiation programs [74]. For example *NFKB2* C-terminus deletions lead to stabilization of the protein and constitutive activation of the NF-κB signaling pathway [73]. Other common mutations in NF-κB signaling pathway in CTCL include TNF Receptor Superfamily

Member 1B (*TNFRSF1B*) activating point mutations and Caspase Recruitment Domain Family Member 11 (*CARD11*) copy number gains or point mutations [66,73,74].

The JAK-STAT signaling pathway is another common constitutively activated pathway in CTCL. Activation of this pathway leads to transcriptional activation of genes associated with proliferation, differentiation and apoptosis. Recurrent gain of function mutations have been identified in *JAK1*, *JAK3*, *STAT3*, and *STAT5B* [66,73]. Gain of function mutations in *STAT3* lead to increased expression of proinflammatory cytokines IL-17 and IL-22 contributing to carcinogenesis [74].

miRNAs are often dysregulated in CTCL. Considering that miRNAs participate in multiple cellular processes, their dysregulation in CTCL can promote or inhibit malignant features [75,76]. One of the most studied miRNAs in CTCL is miR-155, which promotes tumorigenesis. miR-155 downregulates expression of *STAT4* leading to a profile switch from Th1 immune-responsive cytokine production to a Th2 immune-repressive cytokine production [77]. Other miRNAs investigated in CTCL include miR-93, miR-181, miR-21, miR-150, and miR-22 [75].

Considering that CTCL often resembles benign dermatoses, multiple research groups have identified genes that can help distinguish CTCL. Thymocyte selection-associated high mobility group box (TOX) has emerged as a potential biomarker [78]. TOX is a transcription factor that regulates differentiation towards $CD4^+$ lineages but is not expressed in mature T cells [79]. Interestingly, *TOX* expression in CTCL is correlated with thicker lesions, progression and poor prognosis [54]. Other efforts to identify diagnostic biomarkers in CTCL include identification of ectopically expressed genes from developmental programs and stem cell genes [80].

Although recent advances in understanding the molecular characteristics of CTCL have helped bring novel therapies to patients, the molecular understanding of this disease remains limited. Its low incidence, heterogeneity and plasticity represent challenges for a more detailed molecular understanding.

1.4.6 Cell of origin

The cell of origin of a malignancy has great impact in the clinical presentations and molecular features of the disease. However, the cell of origin for CTCL is an active area of research, particularly its clonal origin and the type of cell it arises from. Despite determination of clonality being used as a diagnostic test, clonality is not always identified in patients. Approximately 50% of patients with early disease stages and 100% of patients with late disease stages are positive for a single clone [56]. In response to these discrepancies, recent publications have suggested progression from heterogeneous mutational subclones; these subclones arise from an immature malignant T cell before *TCRB* rearrangement [81-83]. In agreement, it is not clear whether oncogenic mutations arise in the thymus or in mature T cells [66,84].

As for the type of cell CTCL arises, the most widely accepted hypothesis is that this malignancy arises from skin resident memory T cells (T_{RM}). T_{RM} cells are often classified into effector memory cells and central memory cells [85]. The hypothesis of T_{RM} origin is supported by the expression of skin homing receptors such as CCR4 and cutaneous lymphocyte-associated antigen (CLA) [61]. Specifically, it has been hypothesized that MF arises from effector memory T cells, while SS from central memory T cells. SS cells express CCR7, L-selectin, and CD27 suggesting a central memory phenotype and MF cells do not, therefore suggesting an effector memory phenotype [86]. The cell of origin for pcALCL has not been established [64]. Interestingly, the

expression of these lineage specific receptors suggest CTCL malignant cells are mature T cells, in contrast with the hypothesis of origin from an immature malignant T cell.

Normal skin T_{RM} reside in this tissue for defense against viruses, bacteria, fungi and parasites. During an infection, T cells proliferate and expand followed by a contraction in cell number once clearing of the infection is attained. Almost all cells that resulted from the expansion will die by apoptosis, however a few cells become memory T cells. Skin T_{RM} cells express the surface markers CLA, CD69, CD103, among others (**Figure 1.17**). CLA, as mentioned above, binds to E-selectin allowing interaction with skin keratinocytes; CD69 mediates retention in the skin; and CD103 binds E-cadherin to promote retention in epithelial tissues [87,88]. Maintenance of memory T cells in the skin depend on production of IL-15, IL-7 and activated Transforming growth factor beta (TGF-β) by keratinocytes and on the use of lipid metabolism by T_{RM} cells for survival [88].

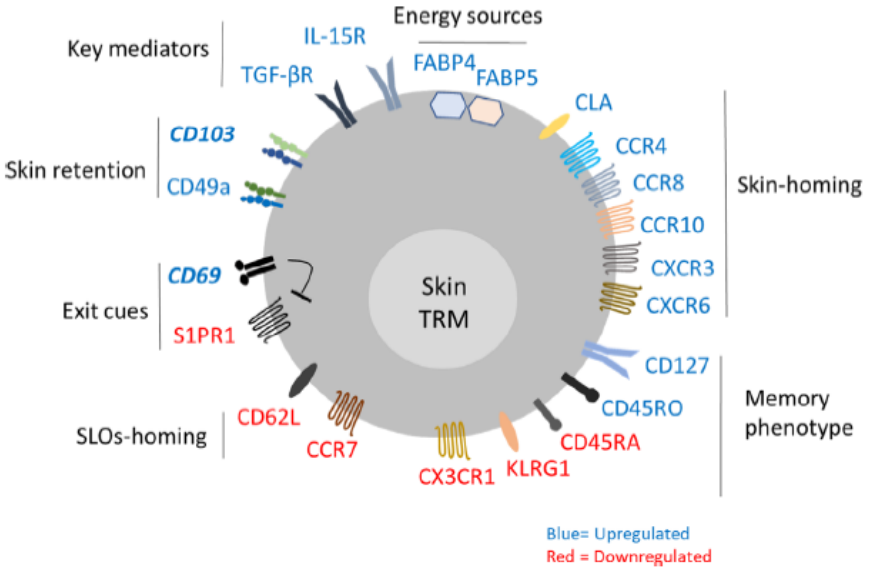


Figure 1.17 Characteristics of skin resident memory T cells.

Upregulation (blue) and downregulation (red) of certain cell surface receptors mediate the characteristics of skin resident memory T cells: skin residency and circulation among other organs and memory phenotype. CCR, C-C chemokine receptor type; CD, cluster of differentiation; CLA, Cutaneous lymphocyte-associated antigen; CXCR, C-X-C Motif

Chemokine Receptor; FABP, Fatty acid-binding protein; IL, Interleukin; KLRG, Killer cell lectin like receptor G; S1PR1, Sphingosine-1-phosphate receptor 1; SLOs, Secondary lymphoid organs; TGF-βR, Transforming growth factor β receptor; TRM, Resident memory T cell. Copied from Figure 1 in [88] used under CC by 4.0.

Another factor to consider in regards of CTCL's cell of origin is phenotypic plasticity. Memory T cells display high plasticity to be able to respond to antigen re-challenges [89,90]; in agreement, CTCL cells display high plasticity. For example, Poglio and colleagues reported that malignant cells when initially isolated from a patient, demonstrated a surface receptor profile of central memory T cells but after culture *in vitro* the cells acquired a naïve phenotype (**Figure 1.18**) [91].

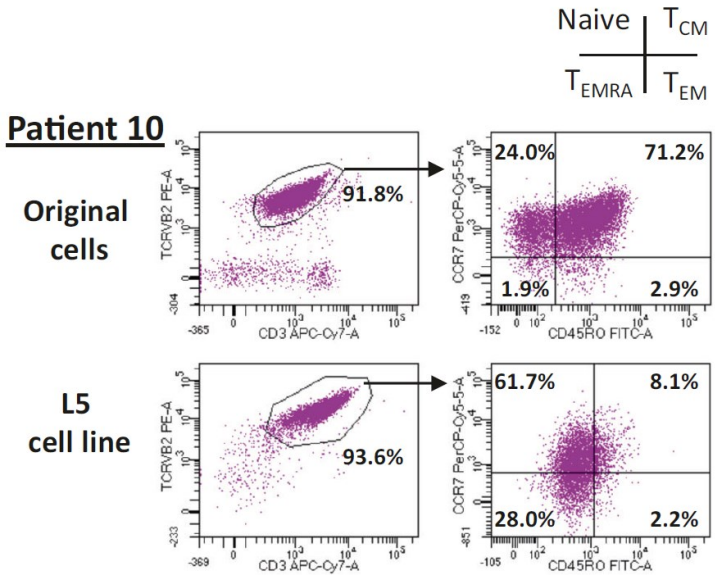


Figure 1.18 Plasticity of CTCL cells.

CTCL primary cells showed a higher percentage of T central memory phenotype characterized by high expression of CCR7 and of CD45RO. After long term culture the phenotype changed to a higher percentage of naïve cells characterized by high expression of CCR7 and low expression of CD45RO. CCR7 PerCRP-Cy5-5-A, C-C chemokine receptor type 7 Peridinin-Chlorophyll-Protein Cyanine5 5 - area; CD, Cluster of differentiation; CD3 APC-Cy7-A, CD3 Allophycocyanin-Cyanine7-area; CD45RO FITC-A, CD45RO+ Fluorescein isothiocyanate-area; L, cell line; TCRVB2 PE-A, T cell receptor variable β 2 phycoerythrin-area; T_{CM}, T cell central memory; T_{EM}, T cell effector memory; T_{EMRA}, effector memory T cells CD45RA+. Adapted from Figure 4B in [91] used under CC by 4.0.

In summary, there is multiple evidence from contrasting hypothesis in regards of the cell of origin in CTCL. Clonality, skin T_{RM} origin and immature T cells are a few hypotheses that have been proposed to understand this disease.

1.4.7 Microenvironment

Establishment of primary cell lines from CTCL samples is challenging, suggesting a high dependency on the tumor microenvironment [91]. CTCL's microenvironment is currently best understood from the cytokines produced by malignant and normal adjacent cells. Early disease stages are associated with production of Th1 cytokines, in particular IFNγ and Tumor necrosis factor (TNF) α. However, malignant cells exhibit a Th2 phenotype with chronic production of IL-4, IL-5 and IL-10 [92]. Disease progression is also associated with a decrease potential of patients to produce Th1 cytokines. Together, these lead to an immune-deficient profile [71].

Other interactions that are relevant to understanding CTCL's microenvironment are with keratinocytes and endothelial cells. It has been shown that production of Th2 cytokines by malignant cells stimulates keratinocytes to produce IL-25. As a consequence, IL-25 further promotes Th2 polarization [93]. Expression of endothelial markers, such as podoplanin in biopsies is associated with disease progression [94]. Langerhans cells, according to their role as APCs, contribute to skin inflammation [95,96].

1.4.8 Hypopigmented Mycosis Fungoides

The following section briefly describes Hypopigmented MF (HMF). Although this is not the main focus of the results presented here, I decided to include a section considering this variant was the initial focus of my PhD studies for two years. The following section is a summary derived from a previously published review in HMF [62].

HMF is a variant of CTCL (**Figure 1.13 B**) characterized mainly by lesions that are light colored or achromic (i.e. loss of skin color). It has been hypothesized that the characteristic presentation of these lesions is caused by damaged and reduced number of melanocytes and/or abnormal melanogenesis [62]. HMF has been reported in African American, South Asian, Middle Eastern and Hispanic individuals, who commonly present darker skin phototypes [62,97]. However, cases of HMF in light skin individuals have also been reported [98]. In contrast to Classic MF, HMF can be diagnosed in pediatric, adolescent and early adulthood populations [62]. However, the prognosis for this variant is better than for Classic MF: The majority of patients are diagnosed at early stages of the disease, and they rarely progress [97]. The last characteristic that differentiates HMF from other variants is the predominance of $CD8^+$ cells [99].

We recently proposed that hypopigmentation of this variant is a surrogate marker of the antitumor immune response in these patients. Particularly, reactive $CD8^+$ T cells secrete toxic granzyme B and granulysin which lead to fewer and damaged melanocytes, to abnormal melanogenesis, and to melanocyte apoptosis. Notably, after treatment HMF patient's skin re-pigments, suggesting the malignant cells are associated with the loss of pigmentation [62].

1.5 Rationale and objective of the research

Based on the background presented above, the current study aims at understanding how reactivation of developmental programs contributes to carcinogenesis. In particular, to understand the role of the germ cell gene GTSF1, a member of the piRNA pathway. Determining the role of GTSF1 in carcinogenesis is clinically relevant due to its potential as an immunotherapy target.

Chapter 2 : Materials and Methods

2.1 Patients and samples

In accordance with the Declaration of Helsinki, all patients enrolled in this study signed an Informed Consent Form. Approval from the ethics review board on each institute involved was obtained prior: The Ottawa Hospital (#20150896-01H), McGill University Health Centre and affiliated hospitals (#A09-M81-10A) and Laval University (# 2011HES-22808). Samples were obtained by punch biopsy and processed into formalin-fixed paraffin-embedded (FFPE) tissue blocks.

2.1.1 Immunohistochemistry

Immunohistochemistry staining of GTSF1 from FFPE blocks was performed by the pathology department of The Ottawa Hospital Research Institute. Staining was performed with the Bond™ III system (Leica) following the standard protocol F. Antigen detection of GTSF1 was performed with the antibody listed in **Table 1** and detected with horseradish peroxidase (HRP) conjugated compact polymer system and 3, 3'-diaminobenzidine (DAB) as chromogen. Slides were counterstained with hematoxylin. Slides were scanned with the Aperio AT Turbo system (Leica). The patient cohort has been previously described [44].

2.2 Analysis of patient data from publicly available databases

RNA-Seq data from 33 cancer types and their normal adjacent tissue (when available) was retrieved from The Cancer Genome Atlas (TCGA) through the cBioPortal (RRID:SCR_014555). The following steps were performed by Dr. Pingxing Xie and Dr. Philippe Lefrançois. Reads Per Kilobase Million (RPKMs) were converted to Transcripts Per Million (TPMs) with the standard formula and mean TPMs of cancer and the normal adjacent tissue were compared with Bayesian analysis with Markov Chain Monte Carlo with the R package rjags (Version 4.3.0

RRID:SCR_017573)[100]. At least 100,000 iterations were performed to estimate p-values. The following steps were performed by me. Multiple hypothesis testing adjustment was performed with R Statistical Software (Version 4.2.1) applying the Bonferroni method. Mean TPMs ± standard deviation (s.d.) were plotted for each cancer type and its normal adjacent tissue.

RNA-Seq raw data in FASTQ format from CTCL patients were retrieved from the NCBI's Gene Expression Omnibus (SRA: SRP309838 and GEO: GSE168508) along with clinical data provided in the supplementary data[101-103]. The following steps were performed by the RNomics Platform at the Université de Sherbrooke. Reads were trimmed using Trimmomatic (Version 0.39 RRID:SCR_011848)[104] and read quality was assessed with FastQC (Version 0.11.9 RRID:SCR_014583)[105]. Read alignment and quantification of transcripts were performed with Kallisto (V0.48.0 RRID:SCR_016582)[106]. Transcriptome of the human genome GRCh38 was created using gffread (cufflinks V2.2.1 RRID:SCR_018965)[107] with the Ensembl annotation and genome files (Version 110). The tximport package (V1.22.0 RRID:SCR_016752)[108] was used to summarize Kallisto count estimates at the gene level. The following steps were performed by me. Patients were ranked based on *GTSF1* mRNA expression level and classified: the upper tertile was classified as high expression and the two lower tertiles as low expression, as previously done[109]. Differential *GTSF1* mRNA expression was evaluated with Mann-Whitney test and survival analysis was performed with Log-rank (Mantel-Cox) test. Box plots for *GTSF1* mRNA expression and Kaplan-Meier plots were graphed.

2.3 Cell culture

All cell lines are human derived. A427, A549, CAL 27, H1299, H23, H28, H460, HuT 78, Kasumi-1, Mac2A, MyLa 2000 (herein referred as MyLa), N/TERT-1, PB2B, UPCI-SCC-090 (herein referred as SCC090) and UPCI-SCC-154 (herein referred as SCC154) are male-derived

cell lines, while A431, Calu-6, HEK293T, H1975 and SZ4 are female-derived cell lines. Mac2A (RRID:CVCL_H637), MyLa (RRID:CVCL_8328), PB2B and SZ4 were obtained from Dr. K. Kaltoft and Dr. N. Ødum (University of Copenhagen, Copenhagen, Denmark). N/TERT-1 (RRID:CVCL_CW92) was obtained from Dr. J. Rheinwald (Harvard Medical School, Boston, USA). A427, A549, H1299, H23, H28, H460 and H1975 were obtained from multiple labs in The Ottawa Hospital Research Institute. HEK293T cells were obtained from Dr. Jean-Jacques Lebrun (RI-MUHC, McGill University, Montreal, Canada). A431 (ATCC Cat# CRL-1555), CAL 27 (ATCC Cat# CRL-2095 RRID:CVCL_1107), Calu-6 (ATCC Cat# HTB-56™ RRID:CVCL_0236), HuT 78 (ATCC Cat# TIB-161™ RRID:CVCL_0337), Kasumi-1 (ATCC Cat# CRL-2724 RRID:CVCL_0589), SCC090 (ATCC Cat# CRL-3239 RRID:CVCL_1899) and SCC154 (ATCC Cat# CRL-3241 RRID:CVCL_2230) were purchased from ATCC. H1975, Kasumi-1, Mac2A, MyLa and PB2B were cultured in RPMI-1640 media (ATCC Cat# 30-2001) containing 10% Fetal Bovine Serum (FBS; Gibco Cat# 12484028) and 1% penicillin-streptomycin (p-s; Gibco Cat# 15140122). SZ4 was cultured in RPMI-1640 media containing 10% FBS and 1% p-s, supplemented with IL-2 (Sigma-Aldrich Cat# I7908-10KU) and IL-4 (Sigma-Aldrich Cat# I4269-5UG) each to a final concentration of 0.001 μg/ml. HuT 78 was cultured in IMDM media (ATCC Cat# 30-2005) containing 20% FBS and 1% p-s. N/TERT-1 was cultured in Keratinocyte-Serum Free Media with Bovine Pituitary Extract (Gibco Cat# 10724-011) and 1% p-s. A431, CAL 27, Calu-6 and HEK293T were cultured in DMEM media (ATCC Cat# 30-2002) containing 10% FBS and 1% p-s. SCC090 and SCC154 were cultured in EMEM media (ATCC Cat# 30-2003) containing 10% FBS, 1% L-glutamine (Gibco Cat# 25030-081) and 1% p-s. Cells were maintained at 37 °C in a humidified incubator containing 5% CO_2.

2.3.1 shRNA-mediated knockdown

To perform short hairpin RNA (shRNA)-mediated GTSF1 silencing, both viral and non-viral methods were employed. For both methods, GIPZ Lentiviral shRNA vectors were purchased from Horizon Discovery. Plasmid details can be found on **Table 2**. For transfection, plasmids were harvested from transformed *E. coli* with CompactPrep Plasmid Preps (Qiagen Cat# 12843 and 12863). Briefly, 0.1 x10^6 H1975 cells/ml were plated and after allowing cells to attach overnight, cells were transfected with 1 µg of plasmid and 4 µL of DharmaFECT (Dharmacon Cat# T-2001-03). Transduction of the three CTCL cell lines was performed with a final concentration of 8 µg/ml of polybrene (Millipore Sigma Cat# TR-1003-G). Mac2a was transduced at Multiplicity of infection (MOI)=5, while MyLa and SZ4 were spinoculated at MOI=0.5 and MOI=10, respectively. Spinoculation was performed at 800 g for 30 minutes at 32 °C. After 48 to 72 hours of transfection or transduction, cells were selected with puromycin for 10 days (H1975 1 µg/ml, Mac2A 1 µg/ml, MyLa 0.5 µg/ml and SZ4 2 µg/ml). To increase efficiency, H1975 single colonies were lifted and expanded in culture. To increase efficiency, CTCL cells were sorted based on Green fluorescent protein (GFP) expression with a BD FACSAria™ Fusion (BD Biosciences) equipped with a yellow-green 561nm laser. Sorting was performed by the Immunophenotyping platform of the RI-MUHC. All cells were maintained in media with puromycin until used for other analyses. For H1975 cell line, clone V3LHS_304725 was used to perform all experiments. For the CTCL cell lines, clone V2LHS_24307 was used to perform all experiments.

2.3.2 Overexpression

For overexpression (OE), Precision LentiORF viral particles were purchased (Dharmacon Cat# OHS5836-EG121355). Transduction was performed following manufacturer's protocol. Briefly, 0.05x10^6 HEK293T cells per well of a 24-well plate were plated and allowed to attach

overnight. Cells were transduced at MOI = 5 with 2 µl of polybrene (Millipore Sigma Cat# TR-1003-G). After 24 hours, antibiotic selection was started with blasticidin at 5 µg/ml (Millipore Cat# 203350). Cells were maintained in media with blasticidin until immunoprecipitation protocol detailed below.

2.3.3 Demethylating agent treatments

Briefly, $0.25x10^6$ cells/ml were plated in T25 flasks in complete media with demethylating agents. Different concentrations of azacitidine (Selleckchem Cat# S1782) and decitabine (Selleckchem Cat# S1200) were tested, ranging from 0.1 µM to 50 µM. After 24 hours, protein was collected as described in the western blot section below. Cells plated in complete media with dimethylsulfoxide (DMSO) were considered the untreated control.

2.3.4 Chemical T cell stimulation

Briefly, $1x10^6$ cells per well of a 24-well plate were incubated in complete media with final concentration of 25 ng/ml of phorbol 12-myristate 13-acetate (PMA; Sigma-Aldrich Cat# P8139-1MG) and 500 ng/ml of ionomycin (Sigma-Aldrich Cat# C7522-1MG). After 6 hours, media was collected by centrifugation at 2,000 x g for 10 minutes at 4°C and stored at -80°C for downstream analysis. Controls for single agent treatment and DMSO were included.

2.3.5 Western blot

Protein was collected as follows. Cells were centrifuged at 2,000 x g at 4°C for 5 minutes and washed twice with cold phosphate buffered saline (PBS; Quality Biological Cat# 119-068-151). Cell lysis was performed with Pierce™ RIPA buffer (Thermo Scientific Cat# 89901) with a tablet of Pierce™ Protease and Phosphatase Inhibitor (Thermo Scientific Cat# A32959). Protein concentration was determined with Pierce™ BCA Protein Assay Kit (Thermo Scientific Cat#

23227). Then, 20-40 µg of total protein were prepared with Sample Buffer (Gen Script Cat# MB01015), boiled at 100 °C for 5 minutes and separated with 4-20% Mini-PROTEAN® TGX Stain-Free™ Precast Gels (Bio-rad Cat# 4568093 and 4568094) with Precision Plus Protein™ ladders All Blue (Bio-rad Cat# 1610373) or Dual Color (Bio-rad Cat# 1610374). Protein was transferred to 0.22 µm polyvinylidene fluoride (PVDF) membranes (Bio-rad Cat# 1704157 and 1704156) with the Trans-Blot® Turbo Transfer System (Bio-rad Cat# 1704150 RRID:SCR_023156). Membranes were blocked for 1 hour with 1% bovine serum albumin (BSA; Sigma-Aldrich Cat# A7906). Membranes were incubated overnight at 4 °C with the appropriate primary antibodies listed in **Table 1**. The next day, membranes were washed with Tris Buffered Saline with Tween 20 (TBST) buffer (150 mM NaCl/20mM Tris base/0.1% Tween 20/pH 7.5) and incubated for 1 hour with the appropriate HRP-conjugated secondary antibody listed in **Table 1**. Membranes were developed with Clarity™ Western ECL Substrate kit (Bio-rad Cat# 1705061) in a ChemiDoc™ MP Imaging System (Bio-rad RRID:SCR_019037). Image processing was done with Image Lab Software (Bio-rad RRID:SCR_014210).

2.3.6 RT-qPCR

RNA was isolated with a RNeasy® Mini Kit (Qiagen Cat# 74104) with DNase treatment (Qiagen Cat# 79254) following manufacturer's protocol. Quantification and quality analysis was performed with a BioDrop µLITE (MBI Cat# 80-3006-55). RNA was converted into complementary DNA (cDNA) with the iScript Advanced cDNA Kit for RT-qPCR (Bio-rad Cat#1725038). Gene expression levels were evaluated with qPCR using SsoAdvanced Universal SYBR® Green Supermix (Bio-rad Cat# 1725274) with the CFX Connect™ Real-Time PCR Detection System (Bio-rad). A list of primers used can be found in **Table 3**. A single housekeeping

gene was used for standardization following the Delta-Delta Ct Method. Mean relative expression from three biological replicates with three technical replicates each ± s.d. were plotted.

2.3.7 MTT assay

Briefly, $0.125x10^6$ cells/ml were plated in a 96-well plate and left overnight to attach. The assay was performed every 24 hours until 168 hours. At each time point, 10 µL of 3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide (MTT; Millipore Cat# 475989) were added to each well and incubated for 2 hours. Then, to stop the reaction 100 µl of 10% sodium dodecyl sulfate (SDS; Sigma Aldrich Cat# L3771) were added and the plate was covered in aluminum foil. The next day, absorbance at 565 nm was recorded with an Infinite® M200 PRO (Tecan RRID:SCR_019033) microplate reader. Absorbance values were corrected with blank wells with media and MTT reagent. Cell numbers were calculated from a standard curve obtained under the same experimental conditions and known cell number. Mean cell number from 3 biological replicates with 24 technical replicates each ± s.d. for each time point were plotted.

2.3.8 Cell proliferation assay

Briefly, $3.8x10^3$ cells per well of a 24-well plate were plated and incubated until each time point. Cell numbers were obtained from 24 to 144 hours with the Vi-CELL XR (Beckman Coulter RRID:SCR_019664) cell counter. Mean cell numbers from three biological replicates with three technical replicates each ± s.d. for each time point were plotted.

2.3.9 Annexin V/PI assay

Briefly, $1x10^6$ cells were collected by centrifugation at 350 x g for 5 minutes followed with PBS washes. Then, cells were resuspended in annexin binding buffer (10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)/150 mM NaCl /2.5 mM $CaCl_2$ in PBS pH 7.4).

Aliquoted cells were stained with propidium iodide (PI; Invitrogen Cat# P1304MP) to a final concentration of 2.5 µg/ml and with 1 µl of annexin V, Alexa Fluor™ 647 conjugate (Invitrogen Cat# A23204) for 10 minutes. Acquisition was performed in a BD FACSCanto II system (BD Biosciences RRID:SCR_018056) with BD FACSDiva Software (Version 8.0.2 BD Biosciences RRID:SCR_001456). FlowJo™ software (Version 10.9.0 BD Life Sciences RRID:SCR_008520) was used for data analysis. The mean percentage from three biological replicates with three technical replicates each ± s.d. of total apoptotic, early apoptotic, and late apoptotic cells were plotted.

2.3.10 Retrotransposition assays

Plasmids for the antibiotic based retrotransposition assay, pJM101/L1.3, pAluA and pCDNA 3.1, were obtained from Dr. John Moran (University of Michigan, Michigan, USA)[110]. The plasmid pUC18 (Cedarlane Cat# SD1162) was used as a negative control. Briefly, $0.06x10^6$ cells/ml in a 6-well plate were plated. After allowing cells to attach overnight, cells were transfected with 9 µL of DharmaFECT (Dharmacon Cat# T-2001-03) and 4 µg of plasmid. After 48 hours of transfection, cells were selected with 500 µg/ml of G418 for 15 days. After selection, wells were washed with PBS, fixed for 1 hour with 4% paraformaldehyde (PFA; Sigma-Aldrich Cat# 158127) and stained for 1 hour with 0.5% crystal violet (Sigma Aldrich Cat# C6158). The assay was repeated with three biological replicates with three technical replicates each. Representative photos of each condition are shown.

Plasmids for the dual luciferase retrotransposition assay, pYX014, pYX015 and pYX017, were obtained from Dr. Wenfeng An (South Dakota State University, South Dakota, USA)[111]. Cells were transiently transfected with the Cell Line Nucleofector® Kit V (Lonza Cat# VCA-1003) and a Lonza™ Nucleofector™ Transfection 2b Device (Lonza RRID:SCR_022262).

Briefly, $2x10^6$ cells were transfected with 20 mg of plasmid and the X-001 program. After 6 hours of transfection, media was changed and cells were incubated for 4 days. Then, cells were collected and lysed according to the Dual-Luciferase® Reporter Assay System (Promega Cat # E1910 and E1980) protocol. Lysates were incubated with the Lysate Assay Buffer II to record Firefly Luciferase values, followed by incubation with the Stop&Glo Buffer to record *Renilla* Luciferase values. Luminescence was recorded with an Infinite® M200 PRO (Tecan RRID:SCR_019033) microplate reader. Each firefly luminescence value was divided by its corresponding *Renilla* value to correct for transfection efficiency and cell survival. Ratios corresponding to pYX015 were used as normalization factors. Mean Relative Luminescence Units (RLU) from three biological replicates with three technical replicates each ± s.d. were graphed.

2.3.11 Immunofluorescence

Adherent and suspension cells were stained with different protocols. Briefly, $0.25x10^6$ adherent cells/ml were plated in 6-well plates with cover slips at the bottom and were allowed to attach overnight. Adhesion of suspension cells was performed as previously described[112]. Briefly, $1x10^6$ cells were resuspended in PBS (Quality Biological Cat# 119-068-151), added to 6-well plates with cover slips at the bottom and allowed to adhere for 30 minutes. Immunostaining of all cells was performed as previously described[113]. Briefly, cells were fixed with 4% PFA (Sigma-Aldrich Cat# 158127), permeabilized with 0.1% Triton X-100 (Bio-Rad Cat# 1610407) and blocked with 1% BSA (Sigma-Aldrich A7906). PBS washes were performed between each step. Incubation with primary antibody listed in **Table 1** was performed overnight at 4 °C in a humid chamber. The next day, cover slips were washed with PBS and incubated for one hour with secondary antibody listed in **Table 1**. Finally, cover slips were washed with PBS and mounted on microscope slides with a drop of Fluoroshield with 4′,6-diamidino-2-phenylindole (DAPI;

Millipore Sigma Cat# F6057). Microscope slides were visualized, and composite photos were acquired with a Lumascope LS720 microscope (Etaluma) equipped with a blue fluorescence filter used for DAPI and a red fluorescence filter used for secondary fluorescent conjugate.

Quantification of Ki67 positive cells was performed with QuPath (version 0.4.4 RRID:SCR_018257) using the Positive cell detection feature with blinded files. A minimum of 500 cells were quantified. Parameters for positive cell threshold were adjusted for each cell line and each marker. For γH2AX and Replication Protein A2 (RPA2) staining, cells were manually counted from a minimum of 25 fields of view. The mean percentage of positive cells from three biological replicates with three technical replicates each ± s.d. were plotted.

2.3.12 ELISA assays

Enzyme Linked Immunosorbent Assay (ELISA) kits to quantify IFNγ (Invitrogen Cat# BMS228), IL-4 (Abcam Cat# ab215089), IL-5 (Abcam Cat# ab215536) and TNFα (Abcam Cat# ab181421) were purchased. Briefly, $1x10^6$ cells per well of a 24-well plate were incubated overnight. Alternatively, cells were chemically stimulated as described above. Supernatant was collected by centrifugation at 2,000 x g for 10 minutes at 4 °C. ELISA assays with supernatants were performed as in manufacturer's protocol, including standard curves. Absorbance values were obtained with an Infinite® M200 PRO (Tecan RRID:SCR_019033) microplate reader at 450 nm. Absorbance values were corrected with the blank wells. Standard curve was plotted with the online Quest Graph™ Four Parameter Logistic (4PL) Curve Calculator (AAT Bioquest https://www.aatbio.com/tools/four-parameter-logistic-4pl-curve-regression-online-calculator) and concentration values were calculated with the equation obtained. For IFNγ and TNFα, mean concentration from three biological replicates with two technical replicates each ± s.d. were plotted. For IL-4, mean concentration from two technical replicates was plotted.

2.3.13 Membrane array

The human Th1/Th2/Th17 Antibody Array kit (Abcam Cat# ab169809) was used. Briefly, $1x10^6$ cells per well of a 24-well plate were incubated overnight. After incubation, supernatant was collected by centrifugation at 2,000 x g for 10 minutes, cleared at 17,200 x g for 10 minutes and immediately tested. Membranes were prepared following manufacturer's protocol. Blocking of the membranes was followed by overnight incubation with cleared supernatant. The next day washes were performed as per the manufacturer's protocol followed by overnight incubation with the biotin-conjugated anti-cytokines. The next day, washes were performed as per the manufacturer's protocol followed by overnight incubation with the HRP-conjugated streptavidin. The next day washes and detection were performed as per the manufacturer's protocol. Detection was performed in a ChemiDoc™ MP Imaging System (Bio-rad RRID:SCR_019037). Quantification was performed with Image Lab Software (Version 6.1 Bio-rad RRID:SCR_014210). Volume values for each dot were obtained from Image Lab, followed by background correction with the negative control values and normalization with the positive control values. Mean volume values from two biological replicates with two technical replicates each ± s.d. for each cytokine tested were plotted.

2.3.14 Flow cytometry of cell surface marker CD25

Briefly, $1x10^7$ cells were collected by centrifugation at 200 x g for 5 minutes, media was discarded, and cells were washed with PBS twice. Cells were aliquoted and stained with 1µl of eBioscience™ Fixable Viability Dye eFluor™ 780 (Invitrogen Cat# 65-0865-14) for 30 minutes at 4 °C. Cells were washed with FACS buffer (2% FBS/1 mM ethylenediaminetetraacetic acid (EDTA) in PBS pH 8.0) and aliquoted to a total of $1x10^6$ cells in a tube with 50 µl of BD Horizon™ Brilliant Stain Buffer (BD Biosciences Cat# 563794). Cells were stained with the fluorochrome-conjugated antibody listed in **Table 1** for 15 minutes in ice. Cells were washed and resuspended

in FACS buffer. Cells were acquired in a BD FACSCanto II system (BD Biosciences RRID:SCR_018056) with BD FACSDiva Software (Version 8.0.2 BD Biosciences RRID:SCR_001456). Data analysis was done on FlowJo™ software (Version 10.9.0 BD Life Sciences RRID:SCR_008520). The mean percentage of positive cells from three biological replicates with three technical replicates ± s.d. were plotted.

2.3.15 Extracellular lactate assay

Lactate production was measured using the Lactate-Glo™ Assay Kit (Promega Cat# J5021). Briefly, $0.2x10^6$ cells/ml were plated in a 24-well plate and incubated overnight. Media was collected and extracellular lactate was measured following manufacturer's protocol. Luminescence was recorded with an Infinite® M200 PRO (Tecan RRID:SCR_019033) microplate reader. Luminescence was normalized to media only as a negative control. Mean RLU from three biological replicates with three technical replicates each ± s.d. were graphed.

2.3.16 Caspase 8 assay

Briefly, $1x10^6$ cells/ml were plated in a 6-well plate and incubated for 10 minutes, 30 minutes, 1 hour, 1.5 hours, 2 hours and 2.5 hours. Caspase 8 measurement was performed with the colorimetric Caspase 8 Assay Kit (Abcam Cat# ab39700) following manufacturer's protocol. After incubation cells were collected, lysed, and protein concentration was determined with Pierce™ BCA Protein Assay Kit (Thermo Scientific Cat# 23227). Volumes were adjusted to 100 µg of protein in 100 µl of cell lysis buffer. Cell lysates were added to 96-well plates with reaction buffer/ Dithiothreitol (DTT) mixture and IETD-*p*NA substrate. Then, the plate was incubated at 37 °C for 1 hour. Absorbance at 400 nm was recorded with an Infinite® M200 PRO (Tecan RRID:SCR_019033) microplate reader. Absorbance was normalized to background wells.

Normalized mean absorbance from one biological replicate with two technical replicates was plotted.

2.3.17 Immunoprecipitation

Immunoprecipitation was performed following manufacturer's protocol of the Dynabeads® Antibody Coupling Kit (Life Technologies Cat# 14311D) or Pierce™ Crosslink Magnetic IP/Co-IP kit (Thermo Scientific Cat# 88805). A total of 40 µg of the antibody against GTSF1 or of normal isotype IgG were coupled to beads following manufacturer's protocol. Antibodies are listed in **Table 1**. Protein was isolated following the steps mentioned in the western blot section above. For immunoprecipitation, 500 µg of total protein was used as initial input. Then, the protein sample was incubated with antibody-coupled beads for 30 minutes at 4 °C in a roller. Flow through was collected. After immunoprecipitation, samples were eluted with PBS and Sample Buffer (Gen Script Cat# MB01015) either at room temperature for 10 minutes or boiled at 100 °C for 5 minutes. Elutes were then analyzed as detailed in the western blot section above, starting with the gel separation. Antibodies used to evaluate precipitation and co-precipitation are listed in **Table 1**.

2.3.18 Mass spectrometry

GTSF1 OE in HEK293T cells was performed as detailed above. Immunoprecipitation with antibody against GTSF1 and normal IgG was performed as detailed in the immunoprecipitation section above for one biological replicate. The following steps were performed by the Proteomics platform at the RI-MUHC. Each sample was processed to remove lipids, detergents, and salts. Then, the sample was reduced with DTT, alkylated with iodoacetic acid, and digested with trypsin. Extracted peptides were re-solubilized in 0.1% aqueous formic acid and loaded onto a Thermo

Acclaim Pepmap (Thermo Scientific 75 µM ID X 2cm C18 3 µM beads Cat# 164946) precolumn and then onto an Acclaim Pepmap Easyspray (Thermo Scientific 75 µM X 15cm with 2 µM C18 beads Cat# ES75500PN) analytical column separation using a Dionex Ultimate 3000 uHPLC (Thermo Scientific) at 250 nl/min with a gradient of 2-35% organic (0.1% formic acid in acetonitrile) over 1 hour. Peptides were analyzed using a Thermo Orbitrap Fusion mass spectrometer (Thermo Scientific) operating at 120,000 resolution (FWHM in MS1) with HCD sequencing (15,000 resolution) at top speed for all peptides with a charge of 2+ or greater. The raw data were converted into *.mgf format (Mascot generic format) for searching using the Mascot 2.6.2 search engine (Matrix Science) against human protein sequences (Uniprot 2021). The database search results were loaded onto Scaffold Q+ Scaffold_5.0.1 (Proteome Sciences) for statistical treatment and data visualization.

2.3.19 RNA-Sequencing

Transcriptomic analysis was performed for the three CTCL cell lines Mac2A, MyLa and SZ4. After knockdown, antibiotic selection and GFP sorting, RNA from three biological replicates of each condition and each cell line was isolated with a RNeasy® Mini Kit (Qiagen Cat# 74104) with DNase treatment (Qiagen Cat# 79254) following manufacturer's protocol. The following steps were performed by Génome Québec. Total RNA was quantified, and its integrity was assessed using 5K/RNA/Charge Variant Assay LabChip and RNA Assay Reagent Kit (Perkin Elmer). Libraries were generated from 250 ng of total RNA as follows: mRNA enrichment was performed using the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs Cat# E7490). cDNA synthesis was achieved with the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs Cat# E7525 and Cat# E7550). The remaining steps of library preparation were done using the NEBNext

Ultra II DNA Library Prep Kit for Illumina (New England BioLabs Cat# E7645). Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the KAPA Library Quantification Kits - Complete kit (Universal) (Kapa Biosystems Cat# KK4824). Average size fragment was determined using a LabChip GX II (PerkinElmer) instrument. The libraries were normalized and pooled and then denatured in 0.02N NaOH and neutralized using HT1 buffer. The pool was loaded at 200pM on an Illumina NovaSeq S4 lane using Xp protocol as per the manufacturer's recommendations. The run was performed for 2x100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1% level. Base calling was performed with RTA (Version 3) and bcl2fastq2 (Version 2.20 RRID:SCR_015058) was used to demultiplex samples and generate FASTQ reads.

2.3.20 Bioinformatic analysis

Downstream analysis of mass spectrometry data was performed by me. The database was cleaned using Scaffold (Version 5.1.2 Proteome Sciences). First, for the proteins with ambiguity, the peptides identified with a probability <90% were unvalidated. Then, the proteins were filtered based on taxonomy (i.e. Homo sapiens) to eliminate contamination from FBS and trypsin. Finally, the database was cleaned based on the CRAPome database [114]. Functional enrichment analysis was performed with g:Profiler with proteins Log_2 fold change ≥ 2 (Version e109_eg56_p17_1d3191d RRID:SCR_006809) with g:SCS threshold method and a significance threshold of 0.05[115,116]. Pathway enrichment graphs were created with GraphPad Prism (Version 10.0.1 RRID:SCR_002798). The interaction map was created with STRING (Version 12.0 RRID:SCR_005223)[117].

RNA-Seq analysis was performed as follows by the RNomics Platform at the Université de Sherbrooke. Raw data were obtained from Génome Québec in FASTQ format. Reads were

trimmed using Trimmomatic (Version 0.39 RRID:SCR_011848)[104] and the quality of the reads was assessed using FastQC (Version 0.11.9 RRID:SCR_014583)[105]. Kallisto (Version 0.48.0 RRID:SCR_016582)[106] was used to align the reads to the transcriptome and to quantify the transcripts. The transcriptome of the human genome GRCh38 was created using gffread (cufflinks Version 2.2.1 RRID:SCR_018965)[107] with the Ensembl annotation and genome files (Version 105). Transcript abundance was combined to obtain the gene level quantification. The tximport package (Version 1.22.0 RRID:SCR_016752)[108] was used to summarize kallisto count estimates at the gene level and DESeq2 (Version 1.34 RRID:SCR_015687) was subsequently used to identify Differentially Expressed Genes (DEGs) between scrambled (SCR) control and GTSF1 knockdown (shGTSF1) conditions, using the default Benjamini and Hochberg correction method. The following steps were performed by me. The functional enrichment analysis was performed with DEGs at Log_2 fold change ≥ 1 and ≤ -1 using g:Profiler (Version e109_eg56_p17_1d3191d RRID:SCR_006809) with g:SCS threshold method and a significance threshold of 0.05 [115,116] following recommendations previously published[118]. Venny 2.1 software was used to identify common DEGs (RRID:SCR_016561). Gene lists used in heatmaps were accessed from KEGG Database (RRID:SCR_012773) [119-121]. Volcano plots, heatmaps and pathway enrichment plots were created with GraphPad Prism (Version 10.0.1 RRID:SCR_002798).

2.4 Statistical analysis

Graphs and statistical analyses were performed using the software GraphPad Prism (Version 10.0.1 RRID:SCR_002798) unless otherwise stated. Normality assumptions were tested with Shapiro-Wilk analysis and the F test. If normality assumptions were met, differences between means of biological replicates were determined by unpaired t test. In the case of the cytokine array, the Mann-Whitney test was used with the Holm-Šídák correction method. In the case of different

γH2AX types of staining comparison was performed with multiple unpaired two-tailed t tests with Welch correction and the two-stage step-up correction method. Error bars represent standard deviation and statistical significance was considered at $p < 0.05$. Figures were created using Inkscape (Version 1.2 RRID:SCR_014479).

2.5 Additional materials and methods

Table 1 Antibodies employed for immunohistochemistry, western blot, immunofluorescence, immunoprecipitation and flow cytometry.

Target	Vendor	Cat#	Application	RRID
β-actin	Cell signaling	4967	WB 1:10,000	AB_330288
γHA2X	Abcam	ab124781	IF 1:1,000	AB_10971675
Anti-mouse IgG HRP linked	Cell Signaling	7076	WB 1:5,000	AB_330924
Anti-rabbit Alexa Fluor 594 Conjugate	Cell Signaling	8889	IF 1:1,000	AB_2716249
Anti-rabbit IgG HRP linked	Cell Signaling	7074	WB 1:5,000	AB_2099233
Cleaved caspase 3	Cell Signaling	9661	WB 1:1,000	AB_2341188
Caspase 3	Santa Cruz Biotechnology	sc-7148	WB 1:200	AB_637828
Cleaved caspase 7	Cell Signaling	9491	WB 1:1,000	AB_2068144
Caspase 7	Cell Signaling	12827	WB 1:1,000	AB_2687912
Caspase 9	Cell Signaling	9502	WB 1:1,000	AB_2068621
Cleaved caspase 9	Cell Signaling	9505	WB 1:1,000	AB_2290727

CD25	Biolegend	302630	Flow cytometry: BV421 conjugated	AB_11126749
DDX4	Abcam	ab27591	WB 1 µg/ml	AB_11139638
GAPDH	Thermo Fisher Scientific	PA1-987	WB 1:2,000	AB_2107311
GMNN	Abcam	ab12147	WB 1:1,000	AB_2110946
GTSF1	Abcam	ab262937	WB 1:1,000 IP 200 µl	-
GTSF1	Abnova	PAB23356	IHC 1:100	AB_11125113
Ki67	Invitrogen	PA5-16785	IF 1:400	AB_11000602
L1 ORF1p	Millipore Sigma	MABC1152	WB 1:1,000	AB_2941775
NFKB2 p100/p52	Cell Signaling	4882	WB 1:1,000	AB_10695537
Normal Rabbit IgG	Cell Signaling	2729	IP 40 µl	AB_1031062
PCNA	Cell Signaling	13110	WB 1:1,000	AB_2636979
PIWIL2	Abcam	ab181340	WB 1 µg/ml	-
PIWIL4	Abcam	ab111714	WB 1:1,000	AB_10887762
p14	Cell Signaling	2407	WB 1:1,000	AB_490785
p16	Cell Signaling	18769	WB 1:1,000	AB_2935679
p21	Santa Cruz Biotechnology	Sc-6246	WB 1:200	AB_628073
pRIPK1 S166	Cell Signaling	65746	WB 1:1,000	AB_2799693
RIPK1	Cell Signaling	4926	WB 1:1,000	AB_2224503
pRb S807/811	Cell Signaling	8516	WB 1:1,000	AB_11178658
Rb	Cell Signaling	9309	WB 1:1,000	AB_823629
RPA2	Invitrogen	MA1-870	IF 1:200	AB_795759
pSTAT3 Y705	Cell Signaling	9145	WB 1:1,000	AB_2491009

STAT3	Cell Signaling	30835	WB 1:1,000	AB_2798995
STAT4	Cell Signaling	2653	WB 1:1,000	AB_2255156
STAT5	Cell Signaling	94205	WB 1:1,000	AB_2737403
pSTAT6 Y641	Cell Signaling	9361	WB 1:1,000	AB_331595
STAT6	Cell Signaling	5397	WB 1:1,000	AB_11220421
S6 ribosomal protein	Cell Signaling	2217	WB 1:1,000	AB_331355
TDRD9	Abcam	ab118427	WB 1 µg/ml	-

Table 2 shRNA plasmids employed in the study

Target	Clone	Method	Cat#	Targeted region
GTSF1	V2LHS_24307	Plasmid	RHS4430-200211965	Non-coding 3' UTR
GTSF1	V3LHS_304723	Plasmid	RHS4430-200265235	ORF
GTSF1	V3LHS_304724	Plasmid	RHS4430-200304589	Non-coding ORF
GTSF1	V3LHS_304725	Plasmid	RHS4430-200302626	Non-coding ORF
GTSF1	V3LHS_304726	Plasmid	RHS4430-200299947	Non-coding ORF
GTSF1	V3LHS_304724	Plasmid	RHS4430-200304589	Non-coding ORF
GTSF1	V2LHS_24307	Viral	VGH5518-200211965	Non-coding 3' UTR
GTSF1	V3LHS_304723	Viral	VGH5518-200265235	ORF
GTSF1	V3LHS_304726	Viral	VGH5518-200299947	Non-coding ORF
Non-silencing	-	Plasmid	RHS4346	Negative control
Non-silencing	-	Viral	RHS4348	Negative control

Table 3 Primer sequences employed in the study

Target	F/R	Sequence
ACTB	F	5’ -CCAACCGCGAGAAGATGA- 3’
ACTB	R	5’ -CCAGAGGCGTACAGGGATAG- 3’
GAPDH	F	5’ -TGATGACATCCAGAAGGTGG- 3’
GAPDH	R	5’ -TTTCTTACTCCTTGGAGGCC- 3’
GTSF1	F	5’ -GCAGACCAGCACCCCATTTGTC- 3’
GTSF1	R	5’ -GGCAGAGATTTGGGAACTCGCA- 3’
L1 ORF1	F	5’ -AGGAAAGCCCATCAGACTAACAGT- 3’
L1 ORF1	R	5’ -GGCCTGGTGGTGACAAAATCT- 3’
L1 ORF2	F	5’ -TCATAAAGCAAGTCCTCAGTGACC- 3’
L1 ORF2	R	5’ -GGGGTGGAGAGTTCTGTAGATGTC- 3’

Chapter 3 : Results – Part I

Ectopic expression of developmental programs in cancer and in CTCL

Contributions: All the work presented in this chapter was contributed by **me**, except: Figure 3.3A, TCGA data retrieval and initial statistical analysis was performed by Dr. Pingxing Xie and Dr. Philippe Lefrançois; figure 3.3C, western blot analysis of GTSF1 expression in a panel of lung cancer cell lines was performed by Dr. Jennifer Gantchev; and figure 3.3E, RT-qPCR of *GTSF1* mRNA expression in a panel of CTCL cell lines was performed by Dr. Jennifer Gantchev. In addition, support for GFP^+ cell sorting was provided by the immunophenotyping platform at the RI-MUHC and mass spectrometry and protein identification was performed by the proteomics platform at the RI-MUHC.

3.1 Preface

Research projects in our lab aim to understand the role reactivation of developmental programs have in carcinogenesis. In particular, we focus on two developmental programs in cutaneous malignancies: TEs and cancer-germline antigens. Some of our work includes reports regarding the reactivation of TEs in CTCL cells and keratinocytes [122,123], and the role the germ cell gene HORMAD1 has in the GIN of SCCs [22].

I am particularly interested in elucidating the role of the germ cell gene GTSF1. GTSF1 is essential for gametogenesis; in mice, GTSF1 controls the expression of TEs [35]. Considering the recurrent reports of the ectopic *GTSF1* mRNA expression in CTCL and its association with a worse prognosis, the investigation of this gene is relevant.

In addition, the expression of these genes in somatic cancer cells can be further exploited for immunotherapy [12,13]. Therefore, there is high relevance and advantages of understanding the role these genes have in carcinogenesis and in the hallmarks of cancer.

3.1.1 Hypothesis

Considering the evidence published from our lab and from other labs, I hypothesized that: The reactivation of TEs in somatic cancer cells induces GTSF1 expression. GTSF1 induces silencing of some TEs via DNA methylation, maintaining a balance of GIN that is beneficial for the cancer cells.

3.1.2 Objective

In chapter 3, I aim to evaluate the reactivation of TEs and the consequences of this reactivation in carcinogenesis, with a focus on the germ cell gene GTSF1.

3.1.3 Experimental Rationale

Through an experimental approach, to determine the role of GTSF1 in carcinogenesis, particularly its impact in GIN. This understanding is clinically relevant. First, *GTSF1* mRNA has the potential to be used as a prognostic tool in CTCL; second, due to its germ cell restricted expression and its presumed immune privilege, GTSF1 can be a valuable prospect for targeted therapy.

3.2 Results

3.2.1 Transposons and their control programs are not constitutively expressed in cell lines representing diverse malignancies

Previous work from our lab evaluated GTSF1 and PIWIL2 protein expression in CTCL cell lines, reporting expression of both at the transcriptional and translational level. [122]. Both proteins are members of the piRNA pathway, therefore, I sought to analyze the expression of other critical members of the piRNA pathway in CTCL and other malignancies, i.e. Tudor Domain Containing 9 (TDRD9), PIWIL4, DEAD-Box Helicase 4 (DDX4) and PIWIL2. I reasoned that a focus on protein expression can provide evidence as to whether the piRNA pathway is activated. Western blot analysis of these proteins in CTCL cell lines revealed a pattern of heterogeneous expression (**Figure 3.1 A**). However, no obvious connection regarding the expression pattern of these different proteins can be discerned. In other words, whether expression of one protein is associated with expression of another one cannot be discerned from this western blot analysis. In

addition, DDX4 is not expressed in any cell line. It has been demonstrated that a lack of DDX4 expression suggests that there is no reactivation of the piRNA pathway in cancer [124]. To evaluate reactivation of TEs, I performed a western blot analysis of L1 ORF1p (**Figure 3.1 B**). ORF1p is one of the two proteins encoded by the only active human retrotransposon L1 [125]. Only the cell line SZ4 demonstrates expression of L1 ORF1p.

Then, to evaluate whether the expression and pattern of expression was specific for CTCL, I performed the same analysis with a panel of cell lines representing other cancer types (**Figure 3.1 C**). All cell lines represent SCCs from different origin organs. Consistently, no clear pattern of expression can be discerned. Furthermore, DDX4 is not expressed in any of these cell lines. Taken together, these analyses suggest neither the piRNA pathway is active in CTCL or SCCs, nor L1 in CTCL.

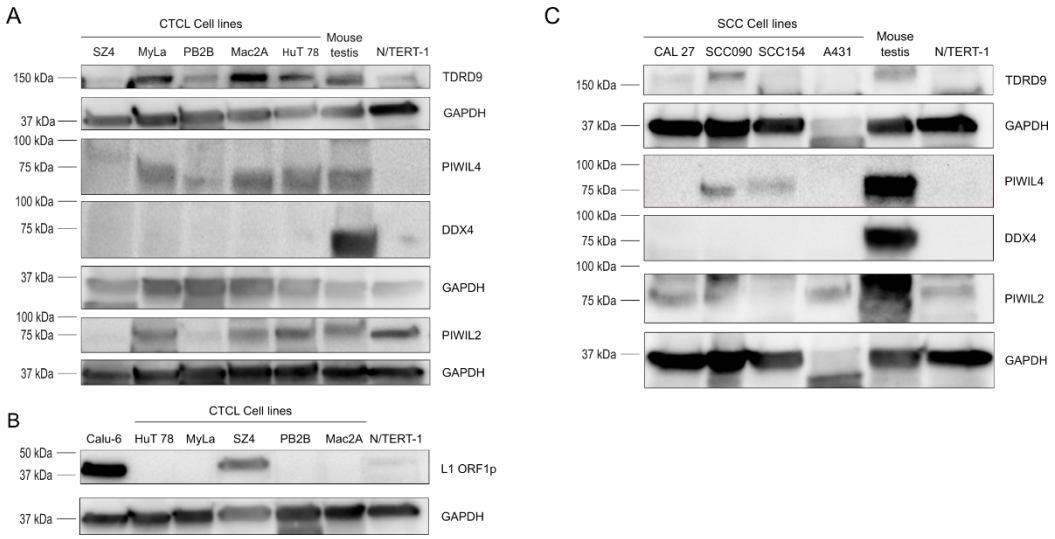


Figure 3.1 Transposons and their control mechanisms are not constitutively expressed in diverse malignancies. **A.** Western blot analysis of piRNA elements TDRD9, PIWIL4, DDX4 and PIWIL2 in CTCL cell lines. Protein from mouse testis is used as a positive control, protein from cell line N/TERT-1 is used as a non-malignant control. GAPDH is used as a loading control and is presented for each membrane probed. **B.** Western blot analysis of L1 ORF1p expression in CTCL cell lines. Protein from cell line Calu-6 is used as a positive control, protein from cell line N/TERT-1 is used as a non-malignant control. GAPDH is used as a loading control. **C.** Western blot analysis of piRNA elements TDRD9, PIWIL4, DDX4 and PIWIL2 in a panel of SCC cell lines. Protein from mouse testis is used as a positive control, protein from cell line N/TERT-1 is used as a non-malignant control. GAPDH is used as a loading control and is presented for each membrane probed.

DNA methylation is one mechanism that cells have developed to control TEs [125]. Therefore, I reasoned that treatment with demethylating agents would reactivate TEs and then, I could evaluate GTSF1 (and other piRNA pathway elements) response. Thus, I treated CTCL cell lines with two demethylating agents, azacitadine and decitabine, for 24 hours (**Figure 3.2 A**). Interestingly, treatment did not lead to a massive reactivation and expression of L1 ORF1p. This suggests a short period demethylating treatment does not lead to strong reactivation of TEs.

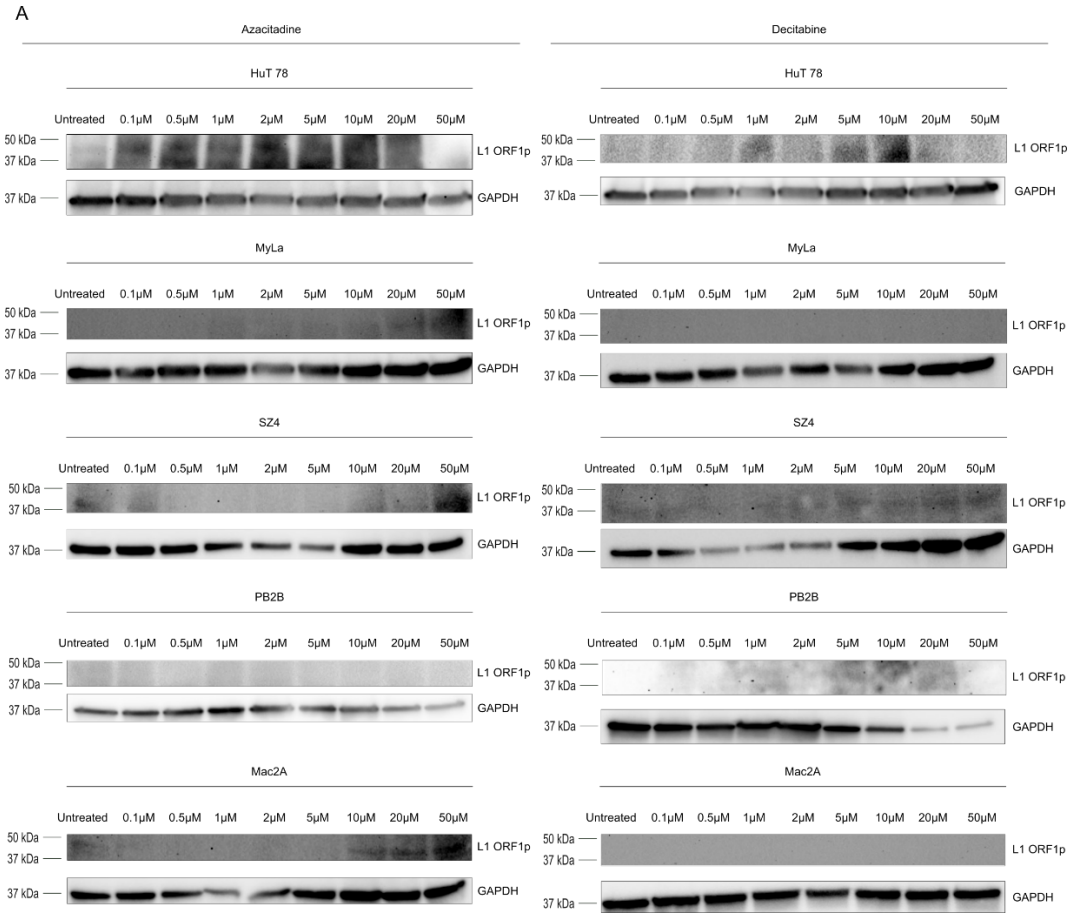


Figure 3.2 Treatment with demethylating drugs do not lead to massive transposon reactivation and expression. **A.** Western blot analysis of L1 ORF1p after 24 hour treatment with demethylating agents with CTCL cell lines. Cells were treated with different concentrations of azacitadine (left) and decitabine (right). Cells incubated with DMSO are considered untreated. GAPDH is used as a loading control.

3.2.2 GTSF1 is ectopically expressed in multiple malignancies

To better understand the expression of *GTSF1* across multiple malignancies, we assessed its transcriptional expression across 33 cancer types with TCGA. Data was retrieved from cBioPortal and statistical analysis was performed. We compared the level of *GTSF1* mRNA expression in cancer and its normal adjacent tissue (**Figure 3.3 A**). This analysis showed that *GTSF1* is significantly overexpressed in Breast Invasive Carcinoma (BRCA), Head and Neck Squamous Cell Carcinoma (HNSC), Kidney Renal Clear Cell Carcinoma (KIRC) and Kidney

Renal Papillary Cell Carcinoma (KIRP). In addition, Acute Myeloid Leukemia (LAML), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) and Testicular Germ Cell Tumors (TGCT) show the highest level of expression, however no normal adjacent tissue data was available to perform statistical analysis. This suggests *GTSF1* is heterogeneously expressed across different cancer types.

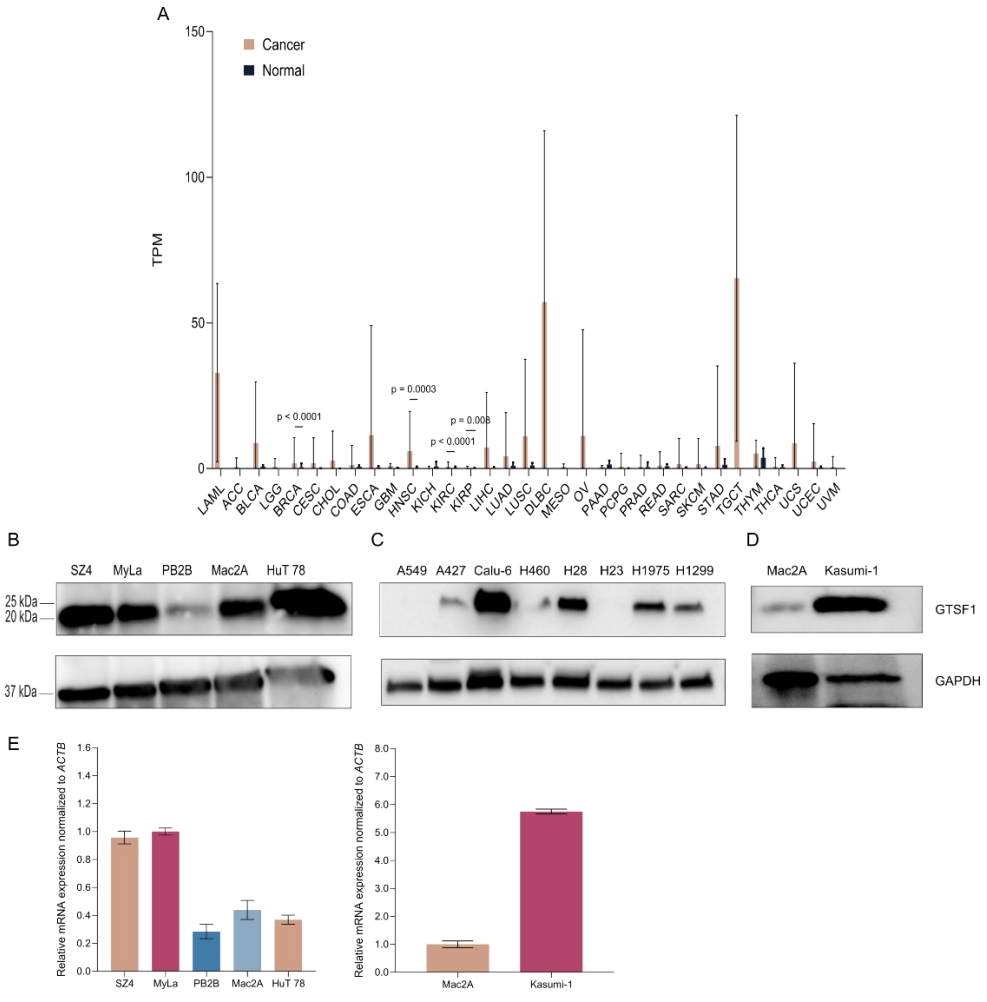


Figure 3.3 GTSF1 is heterogeneously expressed across different cancer types.

A. Relative *GTSF1* expression in Transcript per Million (TPM) across 33 cancer types (brown) and their normal counterpart (black) from TCGA. Differential expression between cancer and normal adjacent tissue was assessed with Markov Chain Monte Carlo method and Bonferroni adjustment with $p<0.05$. Data are presented as means ± s.d. **B.** Western blot analysis of GTSF1 in CTCL cell lines. GAPDH is used as a loading control. **C.** Western blot analysis of GTSF1 in lung cancer cell lines. GAPDH is used as a loading control. **D.** Western blot analysis of GTSF1 in an AML cell line. Protein from cell line Mac2A is used as a positive control. GAPDH is used as a loading control. **E.** Relative

GTSF1 expression normalized to *ACTB* in a panel of CTCL cell lines (left) and an AML cell line (right). Expression is normalized to the highest-expressing cell line (left) or to the lowest-expressing cell line (right) in each graph. Data are presented as means of three biological replicates ± s.d. LAML, Acute Myeloid Leukemia; ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; LGG, Brain Lower Grade Glioma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; COAD, Colon adenocarcinoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; PRAD, Prostate adenocarcinoma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumors; THYM, Thymoma; THCA, Thyroid carcinoma; UCS, Uterine Carcinosarcoma; UCEC, Uterine Corpus Endometrial Carcinoma; UVM, Uveal Melanoma.

To evaluate this heterogenous expression *in vitro*, we performed analyses at the mRNA and protein level. First, we performed western blot analysis with multiple cell lines representing different cancer types. Consistent with previous publications from our lab [42,122], all CTCL cell lines tested showed GTSF1 expression at varying levels (**Figure 3.3 B**). Analysis of a lung cancer cell line panel showed heterogenous GTSF1 expression (**Figure 3.3 C**). Analysis of a single AML cell line also showed high GTSF1 expression (**Figure 3.3 D**). Second, assessment of the mRNA expression showed *GTSF1* expression in CTCL and AML cell lines at varying levels (**Figure 3.3 E**). Collectively, these results suggest GTSF1 expression, both at the mRNA and protein levels, is highly heterogeneous.

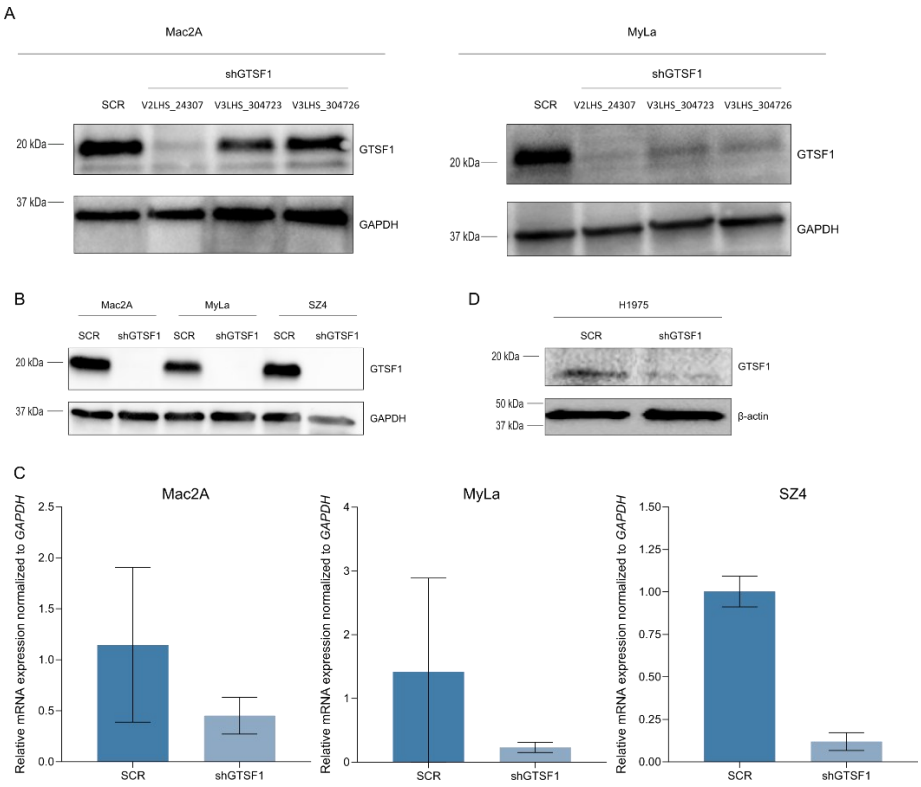


Figure 3.4 Cell line models for studying the role of GTSF1 in carcinogenesis.

A. Western blot analysis of GTSF1 after shRNA knockdown in Mac2A (left) and MyLa (right) with three different plasmid constructs. GAPDH is used as a loading control. **B.** Western blot analysis of GTSF1 after shRNA knockdown in three CTCL cell lines, Mac2A, MyLa and SZ4. GAPDH is used as a loading control **C.** Relative *GTSF1* expression normalized to *GAPDH* after shRNA knockdown in the three CTCL cell lines. Expression is normalized to the non-silencing control (SCR) for each cell line. Data are presented as means of three biological replicates ± s.d. **D.** Western blot analysis of GTSF1 after shRNA knockdown in the lung cancer cell line H1975. β-actin is used as a loading control.

To further understand the role of GTSF1 in carcinogenesis, I developed four cell line models in which GTSF1 expression was silenced with shRNA technology. For studying CTCL, I developed three models. The three selected cell lines represent 80% of all CTCL variants [60]: Mac2A represents pcALCL, MyLa represents MF and SZ4 represents SS [126]. After testing three different shRNA constructs (**Table 2**), I selected the plasmid with the strongest silencing effect at the protein level for each cell line (**Figure 3.4 A**). For these three CTCL cell lines, I used the plasmid V2LHS_24307 to perform all further experiments. The shRNA in this plasmid targets the non-coding 3' Untranslated Region (UTR) (**Figure 3.4 B-C**). For studying other malignancies, I

developed a model with the lung cancer cell line H1975. I used the plasmid V3LHS_304725 to perform all experiments. The shRNA in this plasmid targets the non-coding ORF (**Figure 3.4 D**).

3.2.3 In CTCL and lung cancer, GTSF1 is not a master regulator of proliferation and survival

As an initial approach, I evaluated whether GTSF1 is essential for these cancer cell lines. I hypothesized that the expression of germ cell genes by cancer cells signals a dependency for survival and proliferation. To evaluate proliferation and survival of CTCL cells, I performed a cell proliferation assay from 24 to 144 hours with the three cell lines (**Figure 3.5 A**). Surprisingly, GTSF1 knockdown did not affect cell numbers of any of the cell lines tested at any time point. Due to the limitations associated with the cell proliferation assay, I decided to evaluate proliferation and survival with assays that have higher sensitivity and specificity.

To evaluate whether GTSF1 knockdown led to increased apoptosis, I performed annexin V/PI staining with the three CTCL cell lines. None of the three cell lines showed that GTSF1 knockdown led to an increase in apoptotic cells (**Figure 3.5 B**). Then, to further evaluate the impact of GTSF1 knockdown in proliferation, I performed immunofluorescence staining against the proliferation marker Ki67 (**Figure 3.5 C**). Consistently, Ki67 staining showed GTSF1 knockdown did not impact proliferation of any of the three cell lines.

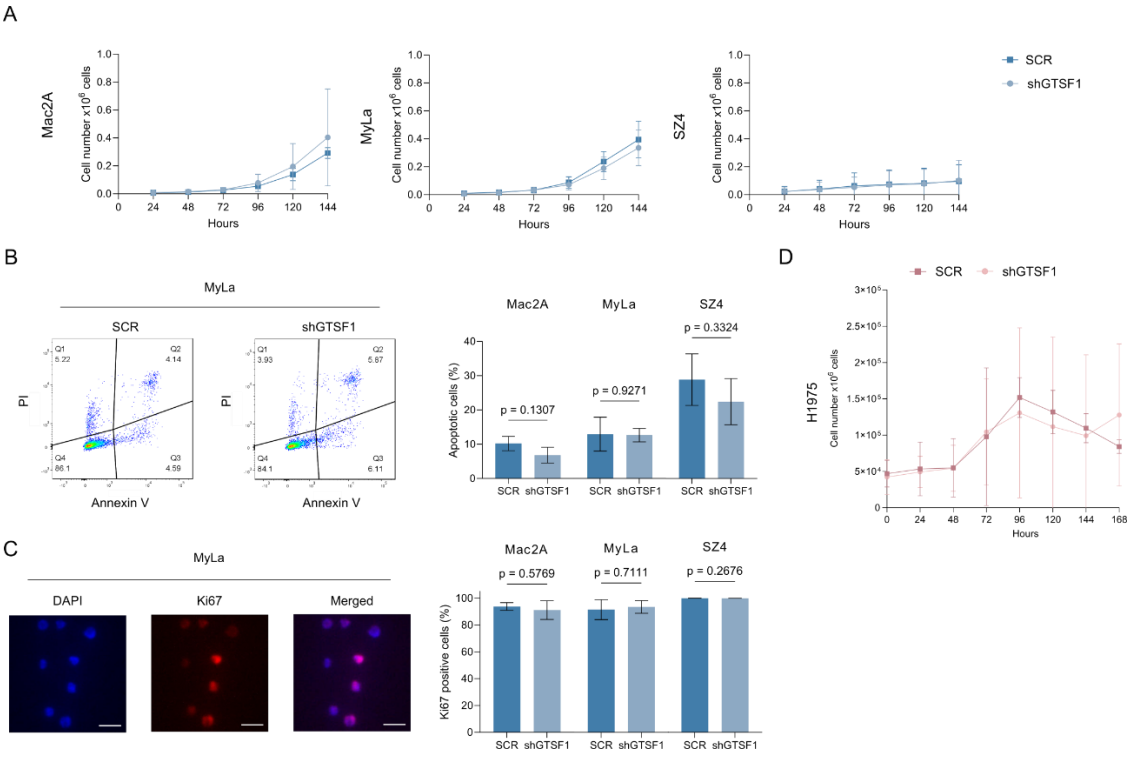


Figure 3.5 GTSF1 is not a master regulator of cell survival in CTCL cells or lung cancer cells.

A. Cell proliferation assay after GTSF1 knockdown for Mac2A, MyLa and SZ4 from 24 hours up to 144 hours. All graphs show SCR (dark blue) and shGTSF1 (light blue). Data are presented as mean cell number of three biological replicates ± s.d. **B.** Annexin V/PI assay after GTSF1 knockdown for Mac2A, MyLa and SZ4. Representative dot plots (left) from MyLa SCR and shGTSF1 are presented. Percentage of total apoptotic cells (right) defined as annexin V^+ + annexin V^+ PI^+ cells. Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three replicates ± s.d. **C.** Ki67 immunofluorescence staining after GTSF1 knockdown for Mac2A, MyLa and SZ4. Representative photos of DAPI, Ki67 and merged channels (left) from Myla. Scale bars represent 20 µm. Percentage of proliferating cells (right) defined as Ki67 positive cells. Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three replicates ± s.d. **D.** MTT assay after GTSF1 knockdown for H1975 from 0 hours up to 168 hours. Graph shows SCR (dark pink) and shGTSF1 (light pink). Data are presented as mean cell number of three biological replicates ± s.d.

To evaluate whether GTSF1 is essential for the cell line H1975, I performed an MTT assay (**Figure 3.5 D**). Consistently, GTSF1 knockdown did not affect cell number at any time point. Taken together, these data suggest that GTSF1 is not a master regulator of survival in CTCL nor lung cancer cells.

3.2.4 In lung cancer, GTSF1 controls transposons

In germ cells from male mice, it has been established that GTSF1 participates in the piRNA pathway [39]. This pathway identifies active TEs and recruits silencing machinery. In a process akin to the maturation of miRNAs, the processing of the piRNAs is aided by GTSF1. Therefore, I evaluated whether in the models I developed, GTSF1 was contributing to control of TEs. Although the piRNA pathway might not be active as suggested above (**Figure 3.1**), other transposon control mechanisms could allow GTSF1 participation. I reasoned that if GTSF1 was controlling TEs expression, its knockdown would lead to an increased expression of TEs and increased transposition events

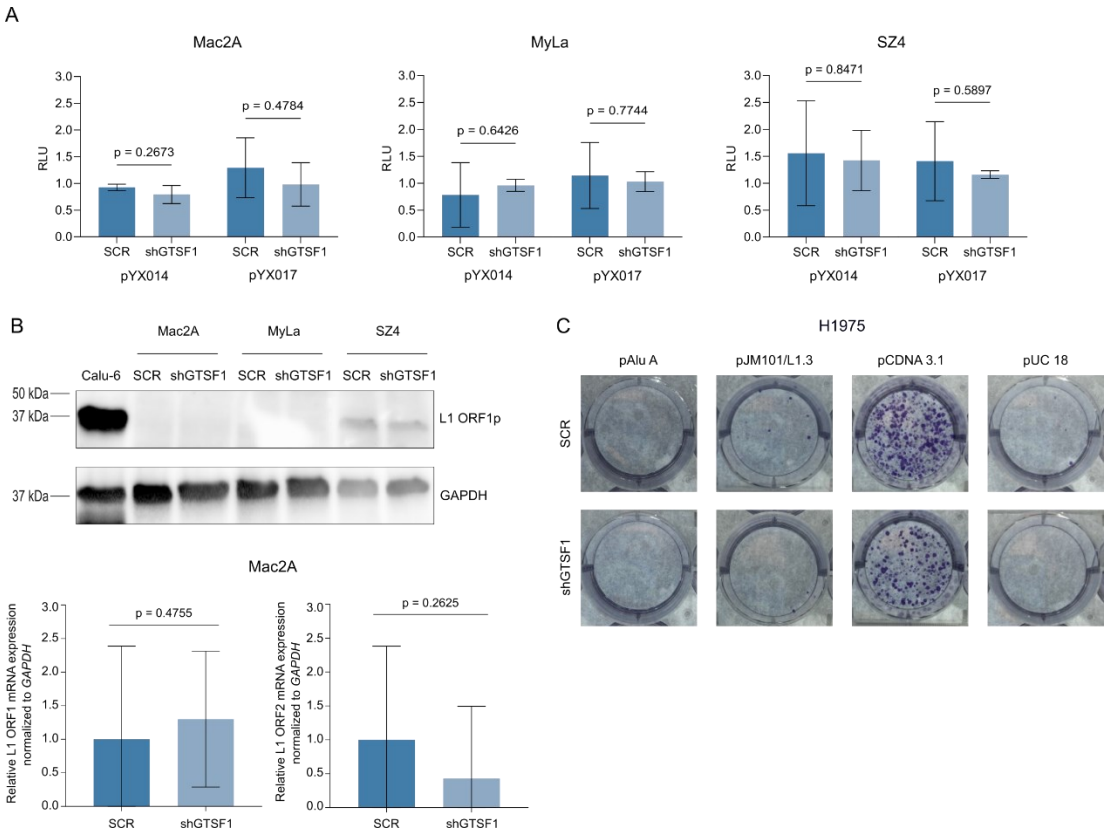


Figure 3.6 GTSF1 modifies transposon activity in lung cancer but not in CTCL.

A. Luciferase activity from dual luciferase retrotransposition assay in Relative Luminescence Units (RLU) after GTSF1 knockdown for Mac2A, MyLa and SZ4. Each assay was performed with two reporter plasmids pYX014 and pYX017

for each CTCL cell line. Differences between SCR (dark blue) and knockdown shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d. **B.** Western blot analysis (top) of L1 ORF1p after GTSF1 knockdown for Mac2A, MyLa and SZ4. Protein from Calu-6 is used as a positive control. GAPDH is used as a loading control. Relative *L1 ORF1* (bottom left) and *L1 ORF2* (bottom right) mRNA expression normalized to *GAPDH* after GTSF1 knockdown for Mac2A. Expression is normalized to SCR. Differences between SCR (dark blue) and shGTSF1 light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d. **C.** Representative photos of antibiotic based retrotransposition assay with two reporter plasmids pAlu and pJM101/l1.3 after GTSF1 knockdown for the cell line H1975. pCDNA 3.1 is used as a positive control and pUC 18 is used as a negative control.

With the CTCL cell line models, I performed a dual luciferase retrotransposition assay [111] (**Figure 3.6 A**). Briefly, only after a full retrotransposition event luciferase would be expressed and luminescence be recorded. I performed this analysis with two different reporter plasmids (pYX014 and pYX017) each with a different promoter. Surprisingly, GTSF1 knockdown did not lead to increased luminescence in any of the three cell lines evaluated. Both plasmids showed consistent results: GTSF1 knockdown did not lead to increased retrotransposition events regardless of the reporter plasmid employed. To further evaluate the potential reactivation of TEs, I evaluated L1 expression at the translation and transcriptional level. Western blot analysis of LI ORF1p showed no increase of expression after GTSF1 knockdown. Consistently, RT-qPCR with Mac2A showed no increased expression of the two L1 ORFs, *ORF1* and *ORF2* (**Figure 3.6 B**).

To evaluate TEs reactivation in the H1975 cell line, I performed an antibiotic based retrotransposition assay [110] (**Figure 3.6 C**). The principle is the same than the dual luciferase retrotransposotion assay: Only after a full retrotransposition event the antibiotic resistance gene will be expressed. I performed this assay with two different reporter plasmids, pAlu A for *trans* retrotransposition and pJM101/L1.3 for *cis* transposition of L1. Interestingly, GTSF1 knockdown decreased the number of colonies obtained after *cis* transposition with pJM101/L1.3; this means that after GTSF1 knockdown fewer cells expressed the antibiotic resistance gene. This suggests

that GTSF1 knockdown led to less transposition events. Taken together, these results suggest that the role of GTSF1 is different in CTCL cells than in lung cancer cells.

3.2.5 GTSF1 knockdown does not increase genomic instability in lung cancer

TEs have multiple mechanisms that allow them to contribute to carcinogenesis [9,10]. One of these mechanisms is facilitating GIN [14]. In our lab we have developed expertise in evaluating this hallmark of cancer [113], therefore I evaluated whether GTSF1 knockdown led to GIN changes in the cell line H1975. Immunofluorescence staining and quantification of γH2AX, a marker of DNA double-strand breaks (DSBs), demonstrated that GTSF1 knockdown did not increase the percentage of DSBs (**Figure 3.7 A left**). An advantage of this staining is that it allows to differentiate levels of GIN: type 1 shows less than 10 nuclear foci indicating low DNA damage; type 2 shows more than 10 nuclear foci indicating high DNA damage; and type 3 shows pan-nuclear staining indicating a pre-apoptotic state [113]. Although I was able to identify all types of γH2AX staining (**Figure 3.7 A middle**), quantification of each type showed no differences after GTSF1 knockdown (**Figure 3.7 A right**). In addition, I performed RPA2 immunofluorescence staining. RPA2, another commonly used marker of DNA damage, binds to single-stranded DNA (ssDNA) [127]. In agreement, GTSF1 knockdown did not increase the percentage of ssDNA (**Figure 3.7 B**). Taken together, these data shows that although GTSF1 knockdown leads to a change in TEs reactivation in H1975, this change does not impact the baseline level of GIN.

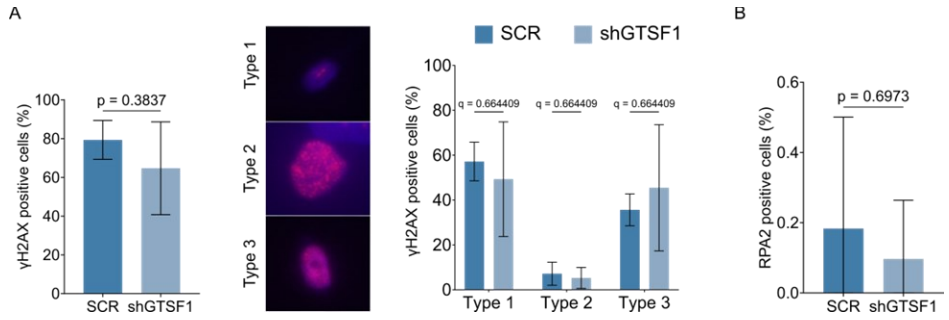


Figure 3.7 GTSF1 knockdown does not increase genomic instability in lung cancer.

A. Immunofluorescence staining for γH2AX after GTSF1 knockdown for H1975. Percentage of γH2AX positive cells (left). Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d. Representative merged photos (middle) of each type of γH2AX staining in H1975. Percentage of γH2AX positive cells with each type of staining (right). Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with multiple unpaired two-tailed t tests with Welch correction and the two-stage step-up correction method. Data are presented as means of three biological replicates ± s.d. **B.** Immunofluorescence staining for RPA2 after GTSF1 knockdown for H1975. Percentage of RPA2 positive cells. Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d.

3.2.6 GTSF1 interacts with protein members of tRNA pathways, DNA synthesis and cellular respiration

In an effort to better understand the context of GTSF1 expression and behavior, I induced OE of GTSF1 in the cell line HEK293T (**Figure 3.8 A**). Following, I performed immunoprecipitation with GTSF1 antibody and normal IgG as an isotype control (**Figure 3.8 B**). Protein interactors were identified with Mass-spectrometry. Fold change analysis between IgG and GTSF1 immunoprecipitation rendered 512 protein interactors with GTSF1. Top protein interactors included DYPSL2, HMGCS1, GMNN, NUP153 and CSNK2A1. To further understand the protein-protein interactions, I created an interaction map with the top 50 enriched proteins identified (**Figure 3.8 C**). Interestingly, GTSF1 (**Figure 3.8 C bottom in light blue**) has no previously reported interaction with any of these proteins. Pathway enrichment analysis of the

interactome, rendered top hit pathways such as cytosolic tRNA aminoacylation, synthesis of DNA and anaerobic respiration (**Figure 3.8 D**).

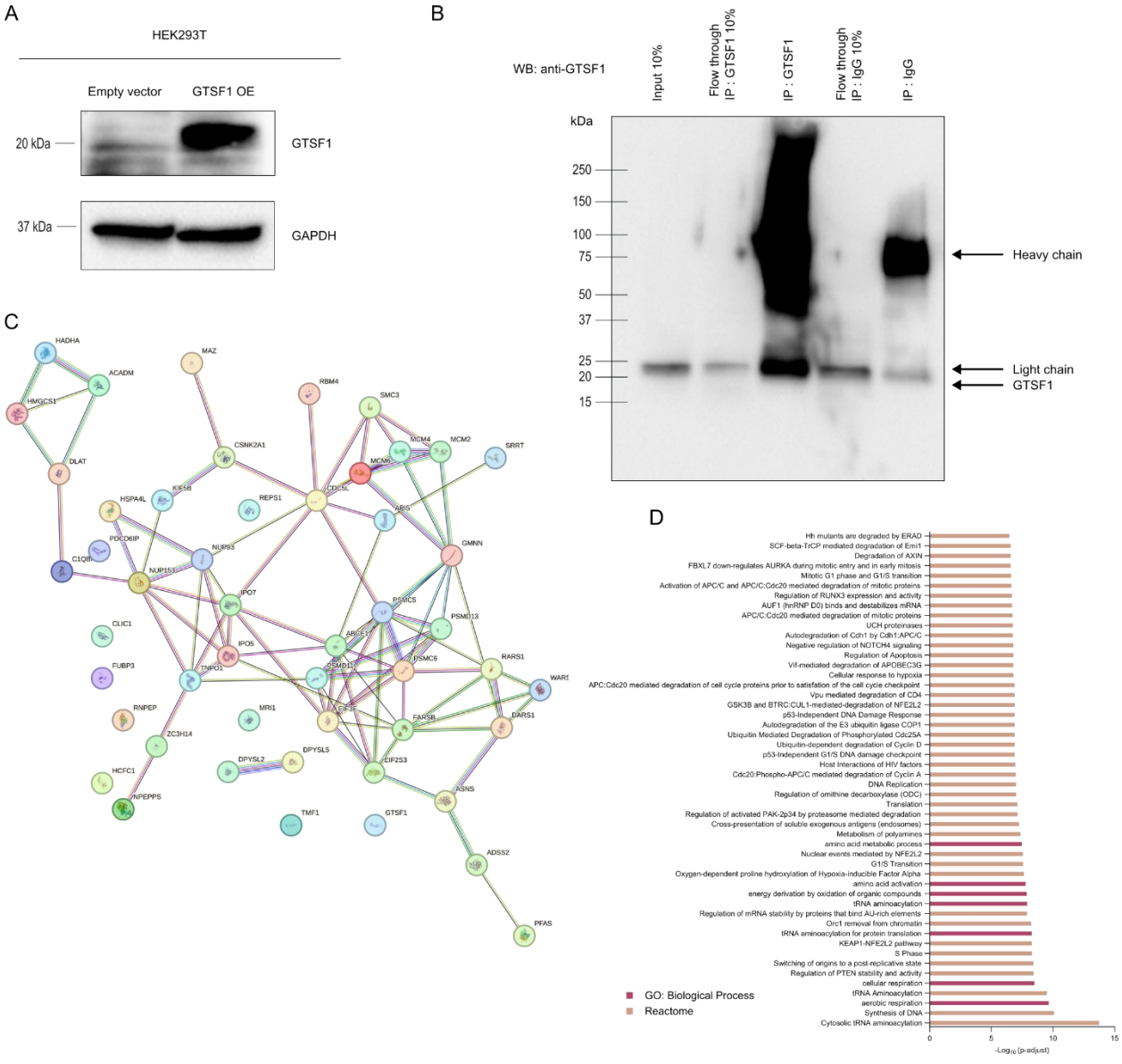


Figure 3.8 GTSF1 interacts with protein members of tRNA pathways, DNA synthesis and cellular respiration. **A.** Western blot analysis of GTSF1 in HEK293T cells after OE. GAPDH is used as a loading control. **B.** Immunoprecipitation with anti-GTSF1 antibody and normal IgG as an isotype control followed by western blot analysis of GTSF1. Lane 1 constitutes 10% of the initial input; lane 2 constitutes the flow through after immunoprecipitation with anti-GTSF1; lane 3 constitutes precipitate with anti-GTSF1; lane 4 constitutes flow through after immunoprecipitation with normal IgG; and lane 5 constitutes precipitate with normal IgG. In lanes 2 and 5 the heavy and light chains of the antibodies can be identified. **C.** Interaction map of the top 50 enriched proteins after anti-GTSF1 immunoprecipitation. Map was created with STRING. GTSF1 can be found at the bottom of the map in a light blue

circle. Mass-spec was performed with one biological replicate. **D.** Pathway enrichment analysis showing -Log_{10}(p-adj) of the top 50 GTSF1 interactors. Pathways from Gene Ontology (GO): Biological Processes are shown in pink and pathways from Reactome are shown in brown. A higher resolution graph can be found in **Appendix 1.**

From the top 50 enriched proteins identified in the OE model (**Figure 3.9 A**), I selected Geminin (GMNN) to validate its interaction with GTSF1 in a CTCL cell line. GMNN is a protein that regulates the cell cycle, and its increased expression has been associated with cancer [128]. I performed anti-GTSF1 immunoprecipitation with lysate from the CTCL cell line Mac2A, followed by western blot analysis of GTSF1 and GMNN. Interestingly, I was not able to detect GMNN co-immunoprecipitation with GTSF1 in Mac2A (**Figure 3.9 B**). This was the only interaction I performed validation.

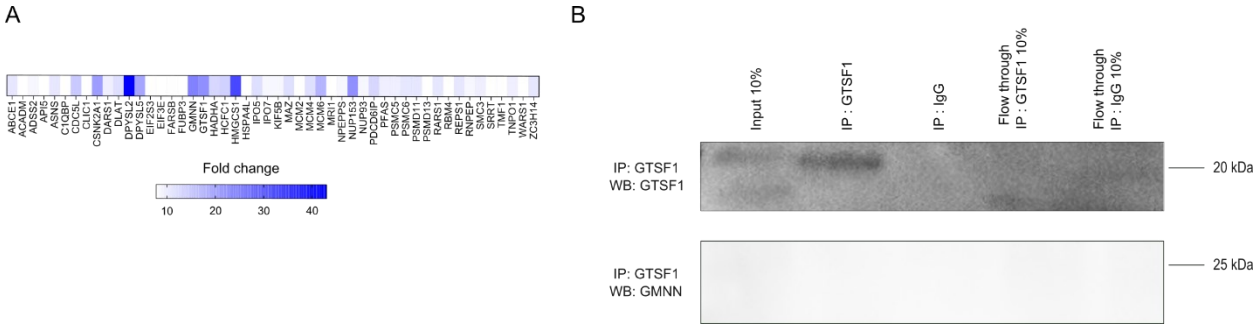


Figure 3.9 GTSF1 interactors in the CTCL cell line Mac2A.

A. Heatmap showing foldchange of anti-GTSF1 immunoprecipitation interactors compared to normal IgG in GTSF1 OE HEK293T cells. Each column represents an interactor and the top 50 interactors are shown in alphabetical order. The foldchange scale is presented at the bottom. **B.** Co- immunoprecipitation with anti-GTSF1 and normal IgG as an isotype control followed by western blot analysis of GTSF1 (top) and GMNN (bottom). Lane 1 constitutes 10% of the initial input; lane 2 constitutes precipitate with anti-GTSF1; lane 3 constitutes precipitate with normal IgG; lane 4 constitutes flow through after immunoprecipitation with anti-GTSF1; and lane 5 constitutes the flow through after immunoprecipitation with normal IgG.

3.3 Chapter 3 conclusions: the role of GTSF1 is context dependent.

In this chapter, I started evaluating reactivation of developmental programs in cancer, narrowing down my research to a specific gene, GTSF1. I have shown that reactivation of TEs and their control mechanisms is heterogeneous and variable in multiple cancer types. Then, I narrowed

my research to GTSF1, a highly relevant gene to the clinical behavior of CTCL. In parallel, evaluation of the same gene in another cancer model allowed me to identify behaviors specific to each cancer type. In lung cancer, GTSF1 modifies the number of transposition events. This suggests that in lung cancer, GTSF1 exhibits its role previously identified in germ cells [39]. In addition, interactome analysis demonstrated that in an OE model, GTSF1 exhibits its binding ability to tRNA [41]. Conversely, the data in this chapter do not demonstrate a specific role of GTSF1 in CTCL. Collectively, these results suggest the role of GTSF1 might be context dependent.

Chapter 4 : Results – Part II

Ectopic expression of germ cell gene GTSF1 in CTCL

Contributions: All the work presented in this chapter was contributed by **me**, except: RNA-Seq processing from cell lines with GTSF1 knockdown and from the CTCL patient database was performed by Dr. Danny Bergeron from the RNomics Platform at the Université de Sherbrooke. Immunohistochemistry of GTSF1 was performed by the pathology department of The Ottawa Hospital Research Institute.

4.1 Bridge between Chapter III and Chapter IV

In CTCL, multiple publications have reported ectopic *GTSF1* expression and its association with a worse prognosis for patients [42-48]. Therefore, this gene presents high clinical relevance.

In the previous chapter I showed that GTSF1 expression in CTCL is not associated with control of TEs. Considering the majority of research in GTSF1 associates it with TEs control, its ectopic expression in CTCL remains an intriguing question. Taking into account these points, I decided to continue my investigation of the role of GTSF1 in carcinogenesis, now focusing only on CTCL.

4.1.1 Hypothesis

Based on the results presented in Chapter III, my refined hypothesis for Chapter IV was: GTSF1 plays a role in carcinogenesis of CTCL.

4.1.2 Objective

In chapter 4, I aim to evaluate the specific role of GTSF1 in CTCL carcinogenesis.

4.2 Results

4.2.1 GTSF1 behaves differently in each CTCL variant

Considering the results presented in Chapter III, I decided to focus on the role of GTSF1 in hallmarks of cancer other than proliferation and growth. Therefore, I decided to employ an unbiased approach with bulk RNA-Seq (**Figure 4.1 A**). We identified the DEGs after GTSF1 knockdown in each CTCL cell line (**Figure 4.1 B**). Surprisingly, each cell line demonstrated a different transcriptomic profile. The number of DEGs identified varied widely, with the most identified in Mac2A, followed by SZ4 and the least identified in MyLa. GTSF1 knockdown rendered 3,954 DEGs in Mac2A, 85 DEGs in MyLa and 1,819 DEGs in SZ4.

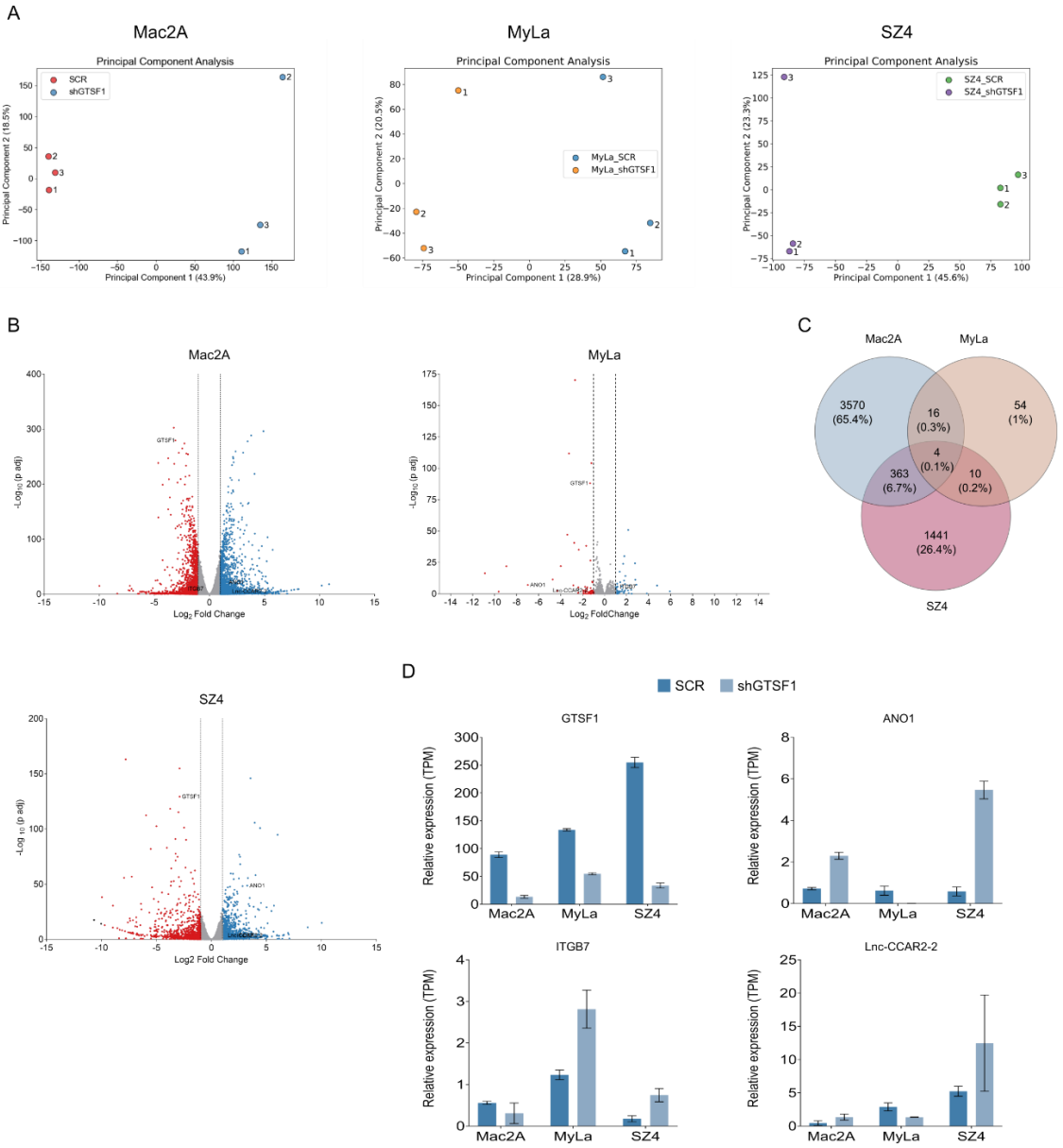


Figure 4.1 GTSF1 knockdown in CTCL variants leads to different transcriptomic profiles.

A. Principal Component Analysis of the normalized RNA-Seq data after GTSF1 knockdown in Mac2A (left), MyLa (middle) and SZ4 (right). **B.** Volcano plots of DEGs showing Log2 fold change and -Log10 (p-adj) after GTSF1 knockdown for Mac2A, MyLa and SZ4. Each dot represents one DEG. Upregulated genes (Log2 fold change ≥ 1) are presented in blue, downregulated genes (Log2 fold change ≤ -1) in red and genes with no significant change in grey. Highlighted are GTSF1, ANO1, ITGB7 and Lnc-CCAR2-2. **C.** Venn diagram of common DEGs after GTSF1 knockdown between Mac2A, MyLa and SZ4. In parenthesis the percentage from the total of queried genes is presented. **D.** Relative expression in TPM of SCR (dark blue) and shGTSF1 (light blue) for the four common DEGs after GTSF1 knockdown between Mac2A, MyLa and SZ4. Data are presented as means ± s.d. Higher resolution of the volcano plots can be found in **Appendix 2, Appendix 3** and **Appendix 4**.

To identify a common role of GTSF1 in the three cell lines, I created a Venn diagram of all the DEGs (**Figure 4.1 C**). This analysis rendered four common DEGs: *GTSF1*, validating the knockdown; anoctamin 1 (*ANO1*), a calcium-activated channel; integrin subunit beta 7 (*ITGB7*), a member of the integrin superfamily; and Lnc-CCAR2-2, a lnRNA without any previously described role. Importantly, except for *GTSF1*, none of these genes presented the same direction of dysregulation (**Figure 4.1 D**). This suggests that after GTSF1 knockdown, each CTCL variant presents a different transcriptomic profile.

To better understand these differences, I decided to evaluate changes in the most common dysregulated pathways in CTCL: TCR, NF-κB and JAK-STAT signaling pathways (**Figure 4.2 A**). Consistently, each cell line presented a different profile. Mac2A presented an overall upregulation of these pathways, MyLa presented both up and downregulation and SZ4 presented downregulation or no changes. To evaluate this at the translational level, I performed western blot analysis of members of these pathways (**Figure 4.2 B**). These analyses confirmed the trend identified in the heatmaps; GTSF1 knockdown in Mac2A led to increased expression and phosphorylation indicating increased activation in these signaling pathways, while SZ4 showed the opposite trend. Interestingly, western blot analysis of NFKB2 (labelled p100 and p52) showed three bands in the Mac2A lane. In CTCL, NFKB2 can present C-terminal deletions which leads to constitutive activation of the NF-κB pathway [129,130]. The identification of an additional band at ~80 kDa suggests Mac2A presents this C-terminal deletion. In addition, Mac2A showed increased phosphorylation of STAT3, suggesting an activation of the STAT3 signaling pathway. MyLa showed minimal changes in the proteins tested. Consistently, SZ4 showed changes in total STAT4 and STAT6, confirming downregulation of these pathways. Taken together, these analyses corroborate a different profile for each cell line representing a different CTCL variant.

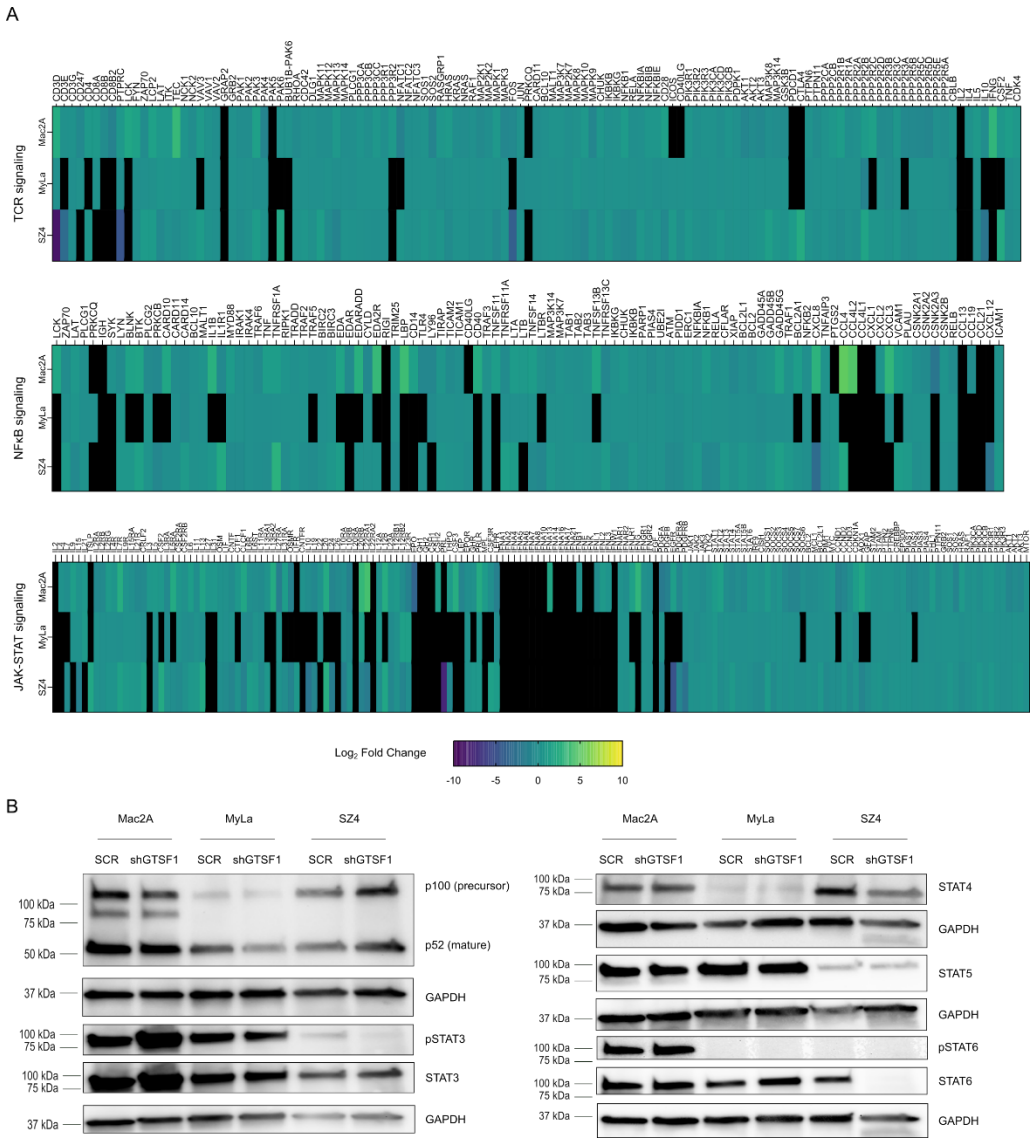


Figure 4.2 Different profile after GTSF1 knockdown in CTCL's most mutated pathways.

A. Heatmaps showing the Log_2 Fold Change of gene expression after GTSF1 knockdown between SCR and shGTSF1 for Mac2A, MyLa and SZ4 for the genes belonging to: TCR signaling pathway (top), NF-κB signaling pathway (middle) and JAK-STAT signaling pathway (bottom). Each column represents a gene and each row a cell line. Gene lists were retrieved from the KEGG database. The Log_2 fold change expression scale is presented at the bottom. A black rectangle in the heatmap cell means no expression was detected. **B.** Western blot analysis of members of the NF-κB and the JAK-STAT signaling pathways: NFKB2 (precursor and mature forms), pSTAT3, STAT3, STAT4, STAT5, pSTAT6 and STAT6, after GTSF1 knockdown in Mac2A, MyLa and SZ4. GAPDH is used as a loading control and is presented for each membrane probed.

CTCL patients often develop immunosuppression, influenced by mutations in the TCR, NF-κB and JAK-STAT signaling pathways [66,71]. Therefore, the reported reactivation of these pathways

in Mac2A represents high clinical relevance. Thus, to better understand this response, I performed pathway enrichment analysis (**Figure 4.3 A**). This analysis showed that GTSF1 knockdown in Mac2A triggers T cell activation. Top hit pathways include: defense response to virus, myeloid leukocyte activation, acute inflammatory response, positive regulation of inflammatory response and positive regulation of tumor necrosis factor superfamily cytokine production. Top hit pathways for SZ4 include: negative regulation of locomotion, integrin cell surface interactions, regulation of chemotaxis and vascular process in circulatory system. MyLa did not render any enriched pathway. Together, these results and their clinical implications, suggested I narrowed my research to the cell line Mac2A.

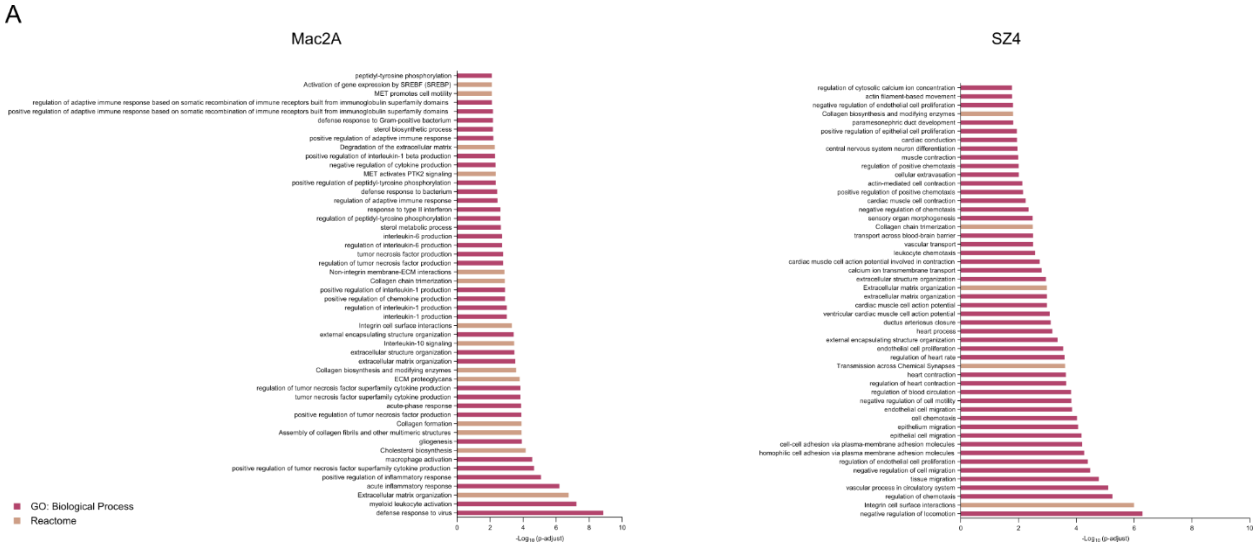


Figure 4.3 Pathway enrichment analysis of DEGs in Mac2A shows T cell activation.

A. Pathway enrichment analysis showing -Log_{10} (p-adj) of DEGs after GTSF1 knockdown for Mac2A (left) and SZ4 (right). MyLa DEGs analysis did not meet the criteria recommended, therefore graph is not shown. Pathways from Gene Ontology (GO): Biological Processes are shown in pink and pathways from Reactome are shown in brown. Higher resolution graphs for Mac2A and SZ4 can be found in **Appendix 5** and **Appendix 6**, respectively.

4.2.2 GTSF1 knockdown leads to T cell activation and cytokine production

Interestingly, one of the top upregulated genes after GTSF1 knockdown in Mac2A was *IFNG* (**Figure 4.4 A top left**). Another mechanism that leads to immunosuppression in CTCL

patients is a shift in the cytokine profile in skin, from a Th1 immune-responsive to a Th2 immune-repressive profile. The Th1 profile is characterized by expression of IFNγ and TNFα, while the Th2 profile is characterized by expression of IL-4 and IL-5 [71]. Moreover, CTCL is thought to arise from skin T_{RM} cells. These T cells exist in a spectrum with effector cells. Therefore, I decided to evaluate other genes associated with the memory/effector spectrum phenotype. GTSF1 knockdown led to expression changes in genes associated with both sides of the spectrum (**Figure 4.4 A**). Genes associated with an effector phenotype, such as *IFNG* and C-X-C Motif Chemokine Receptor (*CXCR*) 3 showed increased expression. At the same time, genes associated with a memory phenotype, such as eomesodermin (*EOMES*) showed increased expression. In addition, genes associated with a Th2 profile, such as *CCR3* and Interleukin 1 receptor-like 1 (*IL1RL1*) showed increased expression. This suggests that GTSF1 knockdown led to a dysregulation in the memory/effector T cell phenotype spectrum.

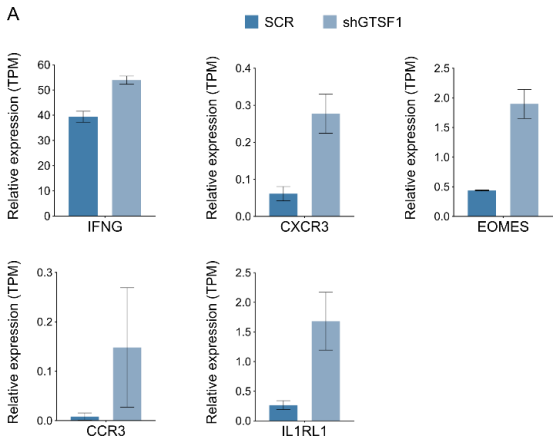


Figure 4.4 After GTSF1 knockdown, Mac2A shows dysregulation in clinically relevant genes.
A. Relative expression in TPM after GTSF1 knockdown of SCR (dark blue) and shGTSF1 (light blue) for clinically relevant genes in Mac2A. Data are presented as means ± s.d.

To further evaluate changes in cytokine production, I evaluated 34 cytokines with a Th1/Th2/Th17 antibody array (**Figure 4.5 A**). The array demonstrated several cytokines were

dysregulated after GTSF1 knockdown. Of particular interest is the increased expression of IFNγ and TNFα. Due to the limitations associated with this analysis, I decided to perform ELISA assays for the clinically relevant cytokines. ELISA assays of the Th1 associated cytokines IFNγ and TNFα, confirmed GTSF1 knockdown led to increased production of these cytokines (**Figure 4.5 B**). By contrast, initial evaluation of the Th2 associated cytokines IL-4 and IL-5 demonstrated no production. To confirm this predominance of Th1 associated cytokines, I chemically stimulated the cells and evaluated IL-4 and IL-5 production. GTSF1 knockdown prevented IL-4 production, only under PMA chemical stimulation (**Figure 4.5 C**). Production of IL-4 was not identified under other stimulation conditions. Taken together, these data suggests that GTSF1 knockdown leads to T cell activation which is accompanied by Th1 cytokine production.

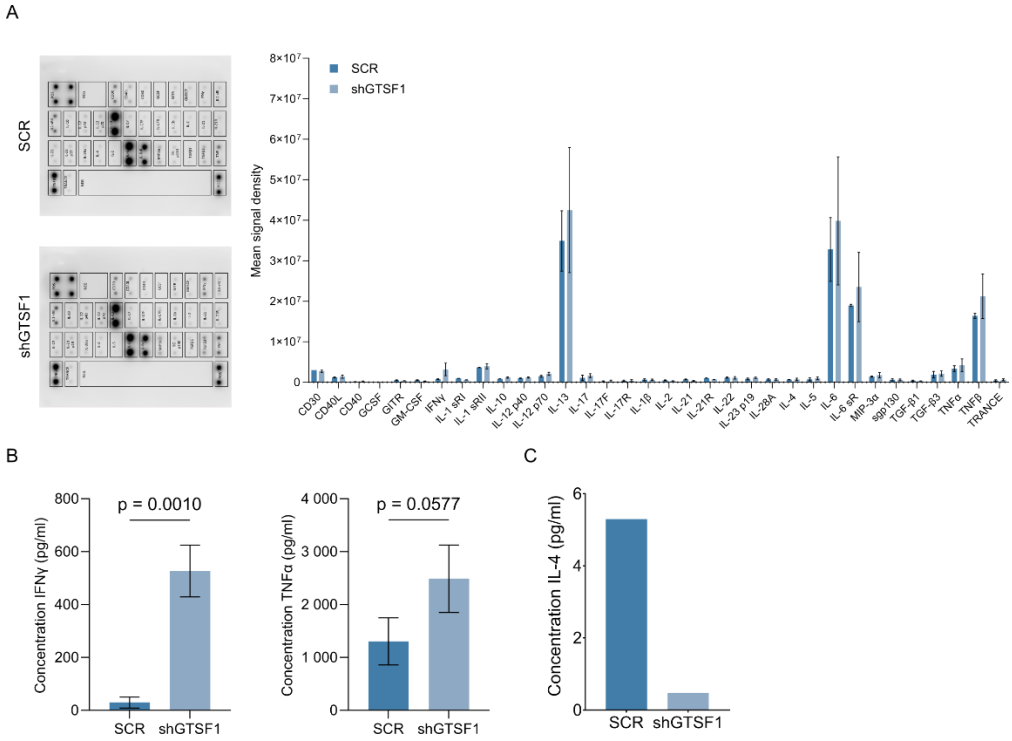


Figure 4.5 Cytokine dysregulation in Mac2A: increased IFNγ and TNFα.

A. Cytokine membrane array analysis after GTSF1 knockdown in Mac2A. Representative images of membranes are shown for SCR (top) and shGTSF1 (bottom). Mean signal density (right) of all cytokines. Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with Mann-Whitney test and Holm-Šídák correction method. Data are presented as means of two replicates ± s.d. All comparisons returned $p \geq 0.999$. **B.** ELISA assays for production of

IFNγ (left) and TNFα (right) from SCR (dark blue) and shGTSF1 (light blue) after GTSF1 knockdown in Mac2A. Differences were evaluated with unpaired two-tailed t test. Data are presented as mean concentration of three biological replicates ± s.d. **C.** Chemical stimulation followed by ELISA assay for production of IL-4 from SCR (dark blue) and shGTSF1 (light blue) after GTSF1 knockdown in Mac2A. Experiment was performed once, and data are presented as mean concentration of three technical replicates.

In addition to cytokine production, T cell activation can be evaluated with other markers. Therefore, I evaluated the surface expression of CD25 after GTSF1 knockdown (**Figure 4.6 A**). CD25 is a well-known marker of T cell activation [85]. Surprisingly, GTSF1 knockdown did not lead to an increased percentage of CD25 positive cells. Another marker of T cell activation is an increased ribosomal and mitochondrial biogenesis [131]. Western blot analysis of the S6 ribosomal protein showed no changes after GTSF1 knockdown (**Figure 4.6 B**). Finally, I evaluated metabolic changes associated with the memory/effector phenotypes. Effector cells favor glycolysis and production of lactate for their metabolic needs, while memory cells favor oxidative phosphorylation [131]. Quantification of the extracellular lactate production showed no changes after GTSF1 knockdown (**Figure 4.6 C**). These contrasting data suggest other markers of T cell activation remain unchanged after GTSF1 knockdown.

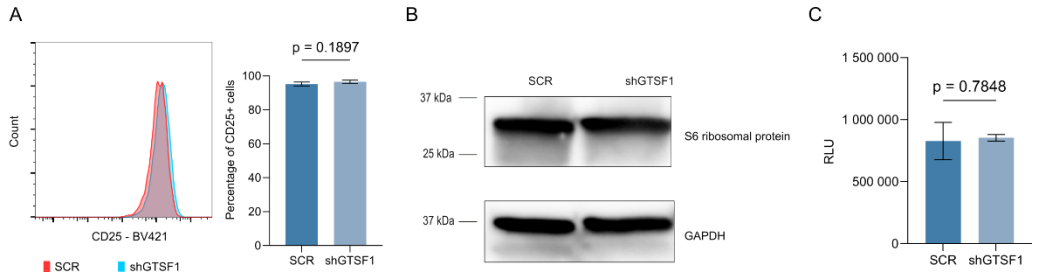


Figure 4.6 Evaluation of additional T cell activation markers.

A. Flow cytometry analysis of cell surface marker CD25. Representative histogram (left) of SCR (red) and shGTSF1 (blue) after GTSF1 knockdown in Mac2A. Percentage of $CD25^{+}$ cells (right). Differences between SCR (dark blue) and shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d. **B.** Western blot analysis of S6 ribosomal protein after GTSF1 knockdown in Mac2A. GAPDH is used as a loading control. **C.** Luciferase-based extracellular lactate assay in Relative Luminescence Units (RLU) after GTSF1 knockdown in Mac2A. Differences between SCR (dark blue) and knockdown shGTSF1 (light blue) were evaluated with unpaired two-tailed t test. Data are presented as means of three biological replicates ± s.d.

4.2.3 Evaluating the consequences of cytokine dysregulation

The contrasting results presented above prompted me to evaluate the consequences of cytokine dysregulation. Some evidence suggests that exposure of T cells to TNF ligands can lead to cell death [132]. Considering that my previous results in Chapter III showed no increased cell death (**Figure 3.5**), I decided to investigate whether the downstream elements of TNF signaling showed any changes. TNF signaling can lead to two possible scenarios, to NF-κB activation or to cell death by apoptosis or necroptosis. Phosphorylation of receptor Interacting Serine/Threonine Kinase 1 (RIPK1) leads to NF-κB activation, survival, and differentiation while lack of phosphorylation of RIPK1 leads to cell death [133]. Western blot analysis at different time points showed GTSF1 knockdown led to lower phosphorylation of RIPK1 levels and higher levels of total RIPK1 (**Figure 4.7 A**). Activity of caspase 8, effector of the apoptosis arm of TNF signaling, was decreased after GTSF1 knockdown (**Figure 4.7 B**). These results are contradictory: A decrease in phosphorylation of RIPK1 suggests that these cells will undergo cell death, however the decreased caspase 8 activity suggests the opposite. Despite this, the involvement of the NF-κB pathway, as shown above (**Figure 4.2**), is further confirmed with these results. Taken together, these data suggest the particular behavior of the NF-κB pathway in Mac2A can be influenced by GTSF1.

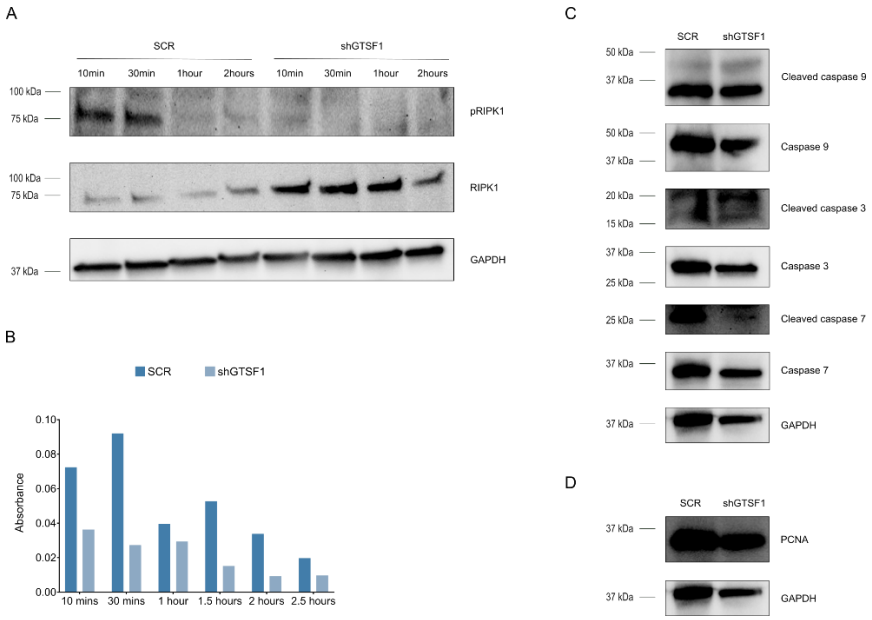


Figure 4.7 Increased TNFα does not trigger TNF-induced cell death.

A. Western blot analysis of members of the TNF-induced cell death pathway, pRIPK1 and RIPK1, after GTSF1 knockdown in Mac2A. Different time points were evaluated. GAPDH is used as a loading control. **B.** Caspase 8 assay after GTSF1 knockdown in Mac2A. Absorbance of SCR (dark blue) and shGTSF1 (light blue) at different time points are shown. Experiment was performed once, and data are presented as average of two technical replicates. **C.** Western blot analysis of total and cleaved caspases: caspase 9, caspase 3 and caspase 7, after GTSF1 knockdown in Mac2A. GAPDH is used as a loading control. **D.** Western blot analysis of the proliferation marker PCNA. GAPDH is used as a loading control. Panels C and D are the same membrane probed for multiple proteins.

To further identify any apoptosis signaling changes after GTSF1 knockdown, I evaluated changes in expression of other members of the apoptotic pathway (**Figure 4.7 C**). Western blot analysis suggests a slight decreased activation of apoptotic pathways after GTSF1 knockdown. However, due to the limits associated with western blot analysis, no definitive conclusions can be drawn. Evaluation of the proliferation marker proliferating cell nuclear antigen (PCNA) further confirmed that increased TNFα production does not lead to changes in proliferation (**Figure 4.7 D**).

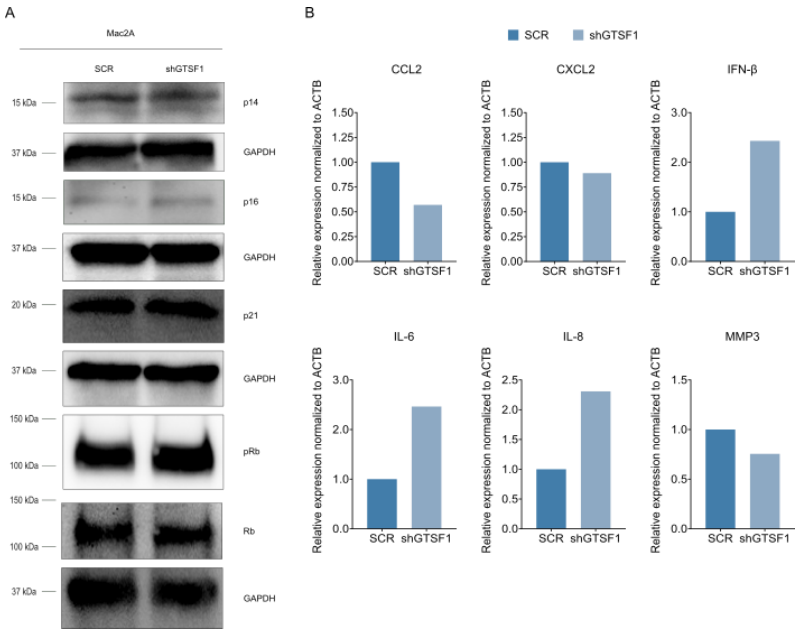


Figure 4.8 Increased TNFα is not associated with cellular senescence.

A. Western blot analysis of cell cycle regulators, p14, p16, p21, pRb and Rb, after GTSF1 knockdown in Mac2A. GAPDH is used as a loading control and is presented for each membrane probed. **B.** Relative expression of SASP genes normalized to *ACTB* after GTSF1 knockdown for Mac2A. Expression is normalized to SCR. Experiment was performed once, and data are presented as mean of three technical replicates.

Another possible explanation for the phenotype reported here is triggering of the senescent program after GTSF1 knockdown. Some characteristics that define senescent T cells are the production of cytokines such as IL-2, IL-6, IL-8, TNF, IFNγ, IL-10 and TGF-β. Production of these cytokines is defined as a senescence-associated secretory phenotype (SASP) [134]. Therefore, I first evaluated changes in cell cycle regulators (**Figure 4.8 A**). Western blot analysis showed no changes in any of the cell cycle regulators evaluated, i.e. p14, p16, p21 and Rb. Then, I evaluated whether GTSF1 knockdown led to expression of SASP genes, *CCL2, CXCL2, IFN-β, IL-6, IL-8* and *MMP3* (**Figure 4.8 B**). GTSF1 knockdown led to changes in the level of expression of these genes, both up and downregulation. These data suggest that GTSF1 knockdown does not lead to senescence in Mac2A cells. Collectively, these data suggest the changes in cytokine production after GTSF1 knockdown influence the NF-κB pathway but do not lead to changes in apoptosis or senescence.

4.2.4 Increased GTSF1 expression is associated with a worse prognosis for CTCL patients

Finally, to evaluate the relationship between *GTSF1* expression and clinical course, we evaluated transcriptomic profiles from CTCL patients. Data was accessed from the publicly available study GSE168508 [101]. Patients were ranked based on the level of *GTSF1* mRNA expression. Based on this rank, I classified the patients in high expression group (highest tertile, n=15) and low expression groups (middle and lowest tertile, n=30) (**Figure 4.9 A**). Interestingly, the two patients with most advanced disease stage (stage IVB), were classified in the high *GTSF1* expression group (**Figure 4.9 B**). Survival analysis demonstrated that patients with high *GTSF1* expression present a worse prognosis: lower overall survival and faster progression (**Figure 4.9 C**). Immunohistochemical analysis of GTSF1 in a previously published [44] patient cohort from our lab confirmed heterogeneous expression in skin samples (**Figure 4.9 D**). Expression can be identified in pleomorphic enlarged cells (**Figure 4.9 D, arrows**)

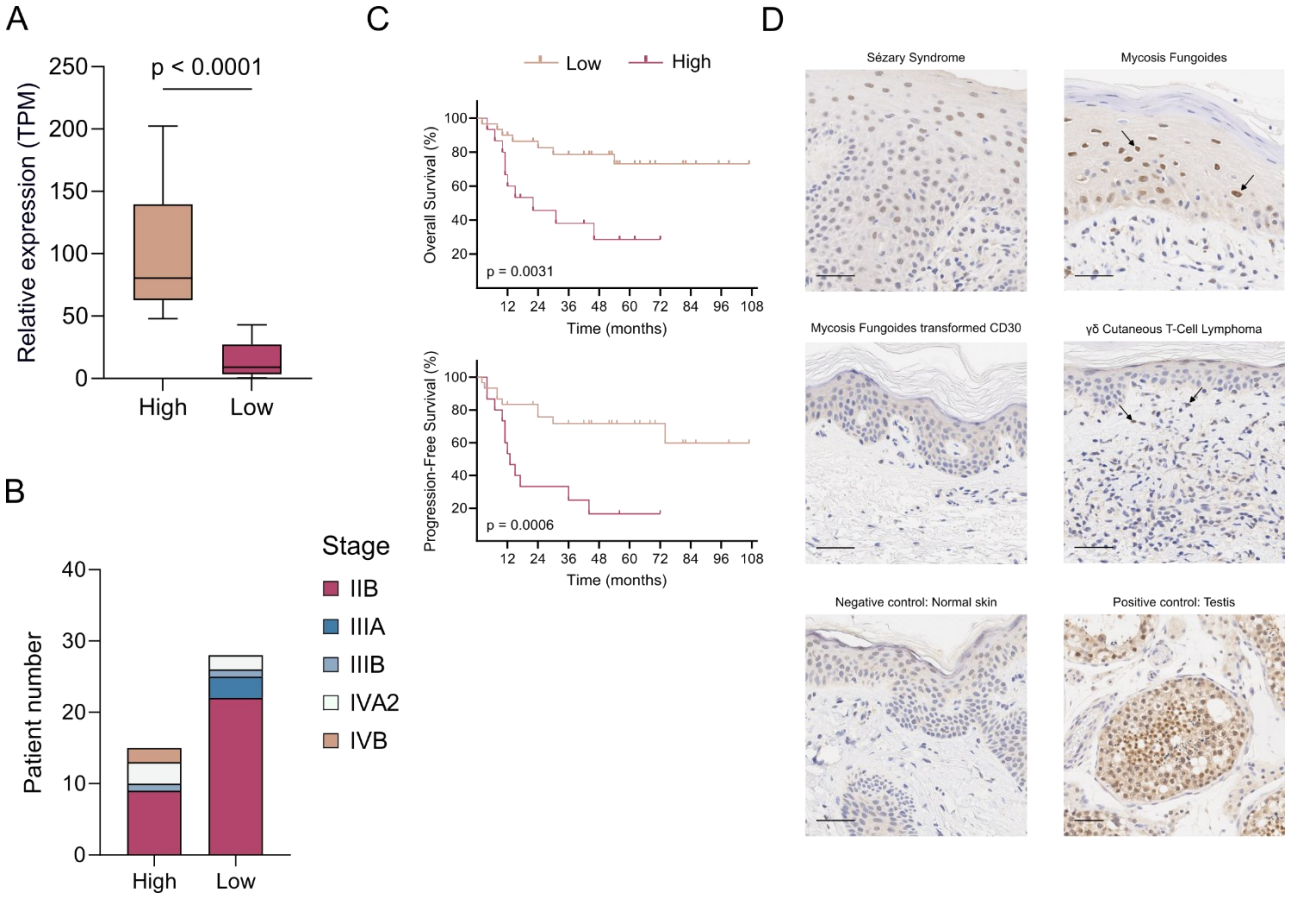


Figure 4.9 Validation of GTSF1 expression in a patient cohort.

A. Relative *GTSF1* expression in TPM of CTCL patients. Differential expression between high (brown) and low (pink) *GTSF1* expression groups was analyzed with Mann-Whitney test with $p<0.05$. Whiskeys represent minimum and maximum and lines in the middle of boxes represent the median. **B.** Number of patients in each disease stage classified by *GTSF1* expression levels. **C.** Kaplan-Meier plots of CTCL patients' disease outcomes: Overall Survival (top) and Progression-Free Survival (bottom). Differences in survival between high (brown) and low (pink) *GTSF1* was identified with Log-rank (Mantel-Cox) test with $p<0.05$. **D.** Representative immunohistochemistry of GTSF1 in skin biopsies from CTCL patients. Each panel represents a different patient. Arrows signal nuclear GTSF1 expression in pleomorphic epidermotropic T cells. Negative control (normal skin) is presented bottom left and positive control (normal human testis) is presented bottom right. Scale bars represent 50 μm.

4.3 Chapter 4 conclusions: GTSF1 modifies the memory/effector phenotype in CTCL

In this chapter, I decided to focus only on the role of GTSF1 in CTCL and the CTCL variant of pcALCL represented by the cell line Mac2a. I have shown that GTSF1 knockdown leads to different profiles for each CTCL cell line, as models of different CTCL variants. GTSF1 knockdown in Mac2A leads to T cell activation and cytokine production. This phenotype suggests

that GTSF1 modifies the memory/effector phenotype of the malignant cells. GTSF1 knockdown led to a partial shift towards the effector phenotype. Interestingly, other markers of effector phenotype, such as CD25 expression or lactate production remain unchanged. In addition, I demonstrated this phenotype is strongly associated with activation of the NF-κB pathway. Finally, I show the potential of *GTSF1* as a prognosis biomarker by analyzing publicly available data from CTCL patients.

Collectively, these results show that GTSF1 plays a role in CTCL carcinogenesis by modifying the production of cytokines. The modification in cytokine production leads to an immune deficient profile in these patients which ultimately allows malignant cells to proliferate and the disease to progress.

Chapter 5 : Discussion

Contributions: All the work presented in this chapter was contributed by **me.**

5.1 Course of the research throughout my PhD

During the first year of my PhD studies, my project was focused on HMF. My supervisor, Dr. Ivan Litvinov, had established a collaboration with Dr. Yann V. Charli Joseph in Mexico City and, considering I lived in Mexico City before my PhD studies, I led this project. HMF is characterized by the loss of skin pigmentation and a good overall prognosis compared to Classic MF. In addition, this variant is more common in Mexico than in Canada, therefore our project aimed at collecting patient samples from Mexico City, snap-freeze them and send them to Montreal for OMICs analyses. We aimed at identifying particular genes that explained the differences between these variants, followed by functional molecular analyses of these genes with wet lab approaches. Even before officially initiating my studies in Montreal, I began working in this project in Mexico City.

Then, at the beginning of my second year of PhD studies, the project had been moving forward slowly. Sample collection had not started, and the panorama was unclear. Therefore, I requested my supervisor to start an additional wet lab project. At around that time, another PhD student at the lab had decided to focus her research project on the reactivation of germ cell genes and GIN. This opened the option for a research project in TEs reactivation and GTSF1 in cancer, therefore I was assigned to this project. Some of the reagents and tools, such as the silencing vectors, were readily available. Considering the challenges of working with suspension cell lines in general, and working with CTCL cell lines in particular, I was advised to begin this project with adherent cell lines. Thus, during the course of my second year of studies I worked with lung cancer cell lines, particularly with the cell line H1975. However, during my second thesis committee meeting, my committee suggested I change the focus of this project to CTCL. They reasoned that because in the initial project submission and first committee meeting, I presented a project with CTCL, I

should continue working with this cancer type. At the same time, I continued to coordinate and manage the collaboration project with three hospitals in Mexico City.

Hence, during the third year of my PhD studies, I started to work on elucidating the role of GTSF1 in CTCL. I continued with this project until the end of my PhD studies. I started working with the five most commonly used CTCL cell lines, but I was advised to reduce this number to three, selecting the ones that would represent the most common variants of CTCL. I standardized the transduction process to obtain multiple cell lines with GTSF1 knockdown. Considering the preliminary results I obtained with the lung cancer cell line, I decided to start this project evaluating TEs reactivation. I decided to evaluate TEs with different experimental approaches that provided more sensitivity. In addition, I decided to select an unbiased approach with RNA-Seq. Interestingly, only the cell line Mac2A showed a clinically relevant phenotype. Consequently, the last year and a half of my PhD studies I focused on deciphering the changes that GTSF1 knockdown triggered in Mac2A.

As for the project in collaboration with Mexico City, I was able to coordinate the delivery of material an equipment necessary for patient sample collection and a few samples were collected. In addition, I established another collaboration for analysis of FFPE HMF samples. Currently, the snap-frozen samples project is cancelled due to unforeseen challenges related to the shipment of biological samples between the two countries. The FFPE samples project is awaiting for the approval of a Material Transfer Agreement between the two institutions.

5.2 Summary and significance of the results presented

Phenotypic plasticity is defined as the ability of cancer cells to dedifferentiate, to block differentiation or to transdifferentiate. It has been proposed that phenotypic plasticity is, at least in

part, enabled by non-mutational epigenetic reprogramming of cancer cells [135]. Interestingly, expression of germ cell genes in cancer is modified by epigenetic reprogramming [11]. Therefore, we can hypothesize that in cancer cells epigenetic changes enable the expression of developmental programs, which in turn unlock further phenotypic plasticity for these cells.

Multiple research groups are actively investigating the specific role ectopically expressed genes have in carcinogenesis. In our lab, we focus on two developmental programs: TEs reactivation and ectopic expression of cancer-germline antigens.

TEs reactivation in cancer has been reported in a plethora of cancer types [136]. However, the exact mechanism in which these elements participate in cancer can vary from cancer type and even from patient to patient. For example, TEs can be inserted in genes responsible for DNA repair, they can cause transcription of oncogenic isoforms, they create DSBs when are activated, among other mechanisms [14].

Ectopic expression of cancer-germline antigens has also been associated with carcinogenesis. For example, we recently reported ectopic expression of HORMAD1 in SCCs regulates their level of GIN. Therefore, HORMAD1 expression is a survival tool for cancer cells, despite presenting high levels of GIN [22].

In addition to the reactivation of developmental programs in cancer, our lab also focuses on elucidating key molecular players in CTCL clinical course. Molecular classification and correct diagnosis remain a challenge in this malignancy [137]. Therefore, previous work in our lab has aimed to identify a gene or a cluster of genes to better diagnose and prognosticate CTCL [42-44,80,138,139]. These publications have demonstrated the utility of evaluating expression changes in cancer-germline antigens to diagnose and/or prognosticate CTCL.

Taking this background, my thesis work had the overarching theme of understanding how the reactivation of developmental programs contributes to carcinogenesis. In particular, my thesis work initially intended to address the knowledge gap regarding the role of GTSF1 in carcinogenesis. In other words, is the ectopic expression of GTSF1 contributing to carcinogenesis? If yes, how is it contributing? Filling this knowledge gap can help us understand why ectopic expression of this gene in CTCL patients is associated with a worse prognosis. Furthermore, filling this knowledge gap can enable the development of better CTCL treatments.

In Chapter 3, I have showed cancer cells acquire expression of developmental programs, such as the piRNA pathway. In germ cells, the piRNA pathway controls the expression of TEs [8]. However, I show that expression of these proteins is heterogeneous and variable across cancer types. A member of the piRNA pathway which shows ectopic expression in cancer is GTSF1. Consistently, I show the potential involvement of GTSF1 in TEs control in the lung cancer cell line H1975. In addition, interactome analysis of GTSF1 suggests an involvement in tRNA aminoacylation; in agreement with a recent publication, GTSF1 interacts with tRNAs to perform TEs silencing [41]. However, in CTCL the evidence did not support a role in TEs control. Together, these results suggest that the role of GTSF1 might be context dependent.

In Chapter 4, I showed the potential role of GTSF1 in modifying the memory/effector phenotype in malignant CTCL cells. This modification is triggered by T cell activation and production of the Th1 cytokines, IFNγ and TNFα. Interestingly, previous publications have associated high expression of *GTSF1* in CTCL patients with worse prognosis [42-48]. However, this association was evaluated with *GTSF1* as part of a gene cluster. I show here that *GTSF1* on its own is associated with worse prognosis. Together, these results suggest GTSF1 expression tilts

the balance of the memory/effector phenotype towards a memory phenotype. Then, the acquisition of a memory phenotype potentially influences the clinical course of the disease.

Therefore, the work presented in this thesis suggests that the ectopic GTSF1 expression does contribute to carcinogenesis. In lung cancer, it suggests participation in TEs reactivation; in CTCL, it suggests modification of the memory/effector phenotype.

5.2.1 Reactivation of developmental programs contribute to carcinogenesis

Research efforts regarding reactivation of developmental programs in cancer hypothesize that they are actively contributing to carcinogenesis. Previous work from our lab reported the ectopic GTSF1 and PIWIL2 protein expression in CTCL cell lines [122]; both proteins participate in the piRNA pathway. In all cell lines I analyzed, key elements of the piRNA pathway do not present an obvious pattern of expression (**Figure 3.1**). Considering that ectopic expression of the germ cell genes in cancer has been associated with epigenetic modifications [11], I treated CTCL cells with demethylating agents. Contrary to what I expected, this treatment did not lead to massive reactivation and expression of TEs (**Figure 3.2**). Taken together, these data suggests that the piRNA pathway might not be active in CTCL or other malignancies.

Other members of the piRNA pathway have been associated with carcinogenesis, in particular PIWIL2 and PIWIL4 [140]. However, the role these publications have reported is not directly associated with the piRNA pathway nor TEs reactivation. For example, it has been reported that PIWIL2 represses p53 by activating STAT3 [141]. Therefore, previous publications support the hypothesis that ectopic expression of these genes is associated with carcinogenesis; however, the evidence suggest that expression does not signifies these genes are performing the same role as in their normal context.

5.2.2 GTSF1 expression in multiple malignancies

In addition to CTCL, ectopic GTSF1 expression has been reported in AML [50,51] and liver cancer [52]. Becker and colleagues evaluated the gene expression signature of AML patients with *WT1* mutations and showed *GTSF1* as the top upregulated gene [50]. Zhang and colleagues identified top miRNAs associated with AML survival. Then, they evaluated the predicted targets of these miRNAs and *GTSF1* mRNA was identified as a top risk factor target gene [51]. Interestingly, our analysis of *GTSF1* expression in cancer and normal adjacent tissue from the TCGA shows high expression of *GTSF1* in AML (**Figure 3.3A, labelled LAML**). In the case of liver cancer, Gao and colleagues reported a high *GTSF1* expression in liver tumor tissues. In their *in vivo* model, GTSF1 OE led to increased tumor development in a mouse xenograft. Conversely, GTSF1 knockdown led to decreased proliferation in hepatoma cell lines [52]. Interestingly, our analysis of *GTSF1* expression in cancer and normal adjacent tissue from the TCGA shows high expression of *GTSF1* in liver cancer too (**Figure 3.3A, labelled LIHC**). However, statistical significance was not reached. Our evaluation of the TCGA data and of multiple cell lines representing CTCL, lung cancer and AML, along with the literature findings, suggest *GTSF1* expression is heterogeneous across cancer types. Unfortunately, none of these publications performed functional analysis. To the best of my knowledge, this thesis and the manuscript associated with it are the first reports of a potential function of GTSF1 in cancer.

5.2.3 GTSF1 is not a master regulator of cell survival and proliferation

It has been hypothesized ectopic expression of germ cell genes provide cancer cells with traits essential for survival or proliferation [2]. Therefore, I hypothesized that if a cancer cell is spending its energy and resources in expressing GTSF1, this might suggest GTSF1 is essential.

Surprisingly, GTSF1 knockdown did not impact cell proliferation or survival in any of the four models (**Figure 3.5**).

Previous publications regarding the reactivation of developmental programs have shown both scenarios: Ectopic expression of germ cell genes can be essential or not for survival and proliferation of cancer cells. For example, ectopically expressed PIWIL2 binds to STAT3 and c-Src. This novel triple protein complex allows c-Src to phosphorylate STAT3, which then represses p53 expression leading to survival and proliferation of cancer cells [33,141]. On the contrary, ectopic expression of preferentially expressed antigen of melanoma (PRAME) in AML induces caspase-independent cell death by decreasing expression of apoptosis inhibiting proteins [142]. Therefore, whilst the results presented here suggest GTSF1 is not a master regulator of cell survival and proliferation, it might contribute to other hallmarks of cancer.

5.2.4 In lung cancer, GTSF1 regulates transposon elements

Previous publications regarding the role of GTSF1 in germ cells reported its participation in the piRNA pathway. This pathway processes piRNAs in a process akin to miRNA processing, termed piRNA biogenesis. Mature silencing complexes translocate to the nucleus to identify active TEs by complementarity and recruit the silencing machinery [8]. Specifically, GTSF1 participates in both steps, grasping and stabilizing piRNA and also interacts with tRNAs and the mature silencing complex to identify their targets [35,41].

TEs reactivation participates in carcinogenesis. The genome-wide demethylation reported in multiple cancer types compromises silencing of TEs [10]. Once reactivated, TEs contribute to carcinogenesis through multiple mechanisms [14]. For example, TEs can cause transcriptional deregulation by inserting in a transcriptional regulatory region or by interacting with

transcriptional regulatory non-coding RNAs. Transcriptional deregulation then can lead to carcinogenesis. TEs can also create GIN through insertional mutagenesis, creating unstable microsatellite seedings, alterations in transcription, creation of DSBs and chromosomal rearrangements [9,10,14].

Considering these points, I hypothesized that GTSF1 knockdown would lead to TEs reactivation (or increased TEs activation from their baseline) and increased GIN. Surprisingly, in the CTCL model GTSF1 knockdown did not increase TEs expression or function (**Figure 3.6A-B**). In contrast, in the lung cancer model GTSF1 knockdown led to decreased transposon events (**Figure 3.6C**) but GIN levels did not change (**Figure 3.7**).

Non-small cell lung cancer, the type H1975 represents, shows a strong connection between hypomethylation of TEs and GIN [143]. Additionally, cancer cells aim to maintain their GIN within certain levels: Cancer cells acquire GIN to adapt but not extreme high levels, so they are able to replicate [144]. Thus, I reason that after GTSF1 knockdown the cells that presented extreme high levels of GIN died by apoptosis. In addition, the antibiotic based retrotansposition assay has low sensitivity which can lead to high levels of false negatives [111]. Moreover, my analysis does not consider the reactivation of TEs other than L1. Therefore, care should be taken when interpreting the results of the retrotransposition assay with H1975. In other words, the evidence presented for the role of GTSF1 controlling TEs in lung cancer needs to be further evaluated. Because my PhD work focused on CTCL, I left many unanswered questions regarding GTSF1 expression in lung cancer.

5.2.5 The role of GTSF1 is context dependent

The results presented here suggest that the role of GTSF1 in carcinogenesis is context dependent. Specifically, my results suggest that in lung cancer GTSF1 participates in TEs control, while in CTCL it participates in regulating the memory/effector phenotype. Discrepancies regarding the role of an ectopically expressed protein in two cancer types have been reported before [140]. For example, in breast cancer PIWIL4 has been implicated in acquisition of mesenchymal characteristics, thereby contributing to the epithelial-to-mesenchymal transition [109]. In contrast, in AML PIWIL4 has been implicated in acquisition of stem cell features and prevention of DNA damage [145]. Therefore, current evidence supports that an ectopically expressed gene can perform different roles depending on the cancer types it is being expressed in. Interestingly, previous publications analyzing GTSF1 in cancer have not focused on TEs reactivation or control.

The current understanding of GTSF1's molecular mechanism suggests that through its interaction with tRNAs, it directs the piRNA pathway to identify active TEs in the genome [35]. Specifically, it is the α-helical portion of the first CHHC zinc finger in GTSF1 which binds to RNA [41]. In alignment with this, interactome and pathway enrichment analyses suggest an interaction with tRNAs (**Figure 3.8** and **Figure 3.9**). Together, these data suggest that GTSF1 is able to participate in sRNA-silencing systems through its RNA interacting domain. Thus, this possible flexibility of GTSF1 for RNA targets contributes to the context dependency hypothesis [146].

5.2.6 The role of GTSF1 is different in each CTCL variant

CTCL is a highly heterogeneous malignancy both at the clinical and molecular level [147-149]. Moreover, CTCL can be classified in multiple variants and subvariants [137]. Here, I selected

three cell lines that would represent one of the three most common CTCL variants, MF, SS and pcALCL; therefore, these three cell lines represent approximately 80% of all CTCL variants [60].

CTCL is highly heterogeneous at the clinical level. The features that define one variant are different than the features that define other variants. For example, some features that define MF include discrete and erythematous lesions in sun-protected skin and $CD4^+$ malignant cells [66]. Meanwhile, SS is an aggressive variant characterized mainly by blood involvement, erythroderma and extreme pruritus [66]. In contrast, pcALCL commonly affects the trunk, face and extremities and lesions can be single or grouped and malignant cells are commonly $CD30^+$ [64]. Furthermore, CTCL is highly heterogeneous at the molecular level. Common mutations are presented in genes associated with DNA damage repair, cell cycle, apoptosis, MAPK pathway, chromatic modifying genes, TCR signaling, JAK-STAT signaling and NF-κB signaling. However, molecular classification has not been possible due to the high heterogeneity that characterizes this malignancy [66]. Therefore, a different role of GTSF1 in each CTCL variant (**Figure 4.1** and **Figure 4.2**) is aligned with the clinical and molecular heterogeneity that characterizes this malignancy.

5.2.7 In pcALCL, ectopic GTSF1 expression modifies the memory/effector phenotype

My investigation suggests that in pcALCL, GTSF1 participates in the memory/effector T cell phenotype. RNA-Seq and cytokine production analysis suggest that GTSF1 knockdown leads to T cell activation and cytokine production (**Figure 4.3**, **Figure 4.4**, **Figure 4.5** and **Figure 4.6**). Current evidence suggests that CTCL arises from skin T_{RM} [86,148], which remain in a resting state until re-challenge with their cognate antigens [85,150]. Interestingly, memory and effector cells are currently understood as two extremes of a spectrum. Resident memory T cells are able to transdifferentiate into effector T cells upon re-challenge [90]. Together, these data suggest that

GTSF1 participates in the mechanisms that drive a cell from an effector towards a memory phenotype.

Some possible explanations for the specificity of this phenotype in pcALCL include mutations in *NFKB2* and the expression of CD30. In Mac2A, western blot analysis (**Figure 4.2B**) showed three bands of NFKB2: p100, the precursor form; p52, the active form; and a band of approximately 80 kDa. It has been previously reported a C-terminal deletion in *NFKB2* rendering the NF-κB pathway constitutively active [129,130]. Another possible explanation is the influence of CD30 in the behavior of this variant. In pcALCL, high expression of CD30 has been associated with activation of the NF-κB and MAPK pathways that lead to cell proliferation [151]. Taken together, current evidence suggests the NF-κB pathway is playing an essential role in the phenotype of Mac2A after GTSF1 knockdown. Evaluation of these possible explanations can help identify the specific context in which GTSF1 modifies the memory/effector phenotype.

5.2.8 The role of the NF-κB signaling pathway in T cells and CTCL

The NF-κB signaling pathway is a major regulator of immune responses. The mechanisms in which this pathway modifies the T cell lineage depend largely on the stage or cell type [152]. In mature T cells, TCR signaling activates the NF-κB pathway through a kinase signaling cascade and recruitment of a signaling adaptor network. This network then activates the canonical and the noncanonical NF-κB pathway [152,153]. Nonetheless, the specific role of NF-κB in memory T cells has been challenging to dissect [153].

The current evidence presented here suggests that the NF-κB pathway is playing an essential role in the phenotype of Mac2A after GTSF1 knockdown. Multiple studies have shown that in CTCL this pathway is constitutively activated [66]. Commonly mutated members of this

pathway include: point mutations or amplifications of TNF receptor 2 (*TNFR2*); copy number gain and activating point mutations in *CARD11*; and C-terminus deletions in *NFKB2* [154-157]. In pcALCL the expression of CD30 further involves this pathway [151]. A recent model, suggests that STAT3 mutations induce its transcriptional activity, leading to expression of NFKB2 and CD30, then, expression of CD30 induces constitutive NF-κB activation [158]. Accordingly, GTSF1 knockdown led to increased phosphorylation of STAT3 (**Figure 4.2B**). Taken together, the increased activation of STAT3 and the potential C-terminus deletion in *NFKB2* discussed above (**Figure 4.2** and **5.2.7**) suggests the NF-κB pathway is playing an essential role in the phenotype of Mac2A after GTSF1 knockdown.

5.2.9 Increased production of clinically relevant cytokines

In CTCL, disease progression is associated with a shift in the type of cytokines produced in the skin. In early disease stages, skin samples show high expression of the Th1 immune-responsive cytokines IFNγ and TNFα; In the late disease stages, skin samples show high expression of the Th2 immune-repressive cytokines IL-4 and IL-5 [46,48,92,159].

Skin T_{RM} produce cytokines, such as IFNγ and TNFα, under inflammatory conditions or stimulation [87]. Here, I demonstrate that GTSF1 knockdown led to an increase in production of the Th1 immune-responsive cytokines IFNγ and TNFα (**Figure 4.5B**). Previous work demonstrated that even after T cell stimulation, the cell lines employed here do not produce IFNγ [77]. In addition, chemical T cell stimulation with PMA demonstrated that GTSF1 knockdown decreased the potential to produce the Th2 immune-repressive cytokine IL-4 (**Figure 4.5C**). Intriguingly, stimulation with ionomycin or with PMA + ionomycin did not lead to changes in IL-4 production. PMA directly stimulates protein kinase C, downstream of the TCR [160], suggesting the relevance of these protein in GTSF1 knockdown phenotype. Taken together, these suggest that GTSF1

knockdown triggered a response similar to when T_{RM} cells are under inflammatory conditions or stimulation, with a preference to produce Th1 cytokines. Furthermore, this is aligned with the ability of memory T cells to transdifferentiate into effector T cells upon re-challenge [90]. In other words, these suggest that through its participation in the mechanisms that drive a cell towards a memory phenotype and the consequent decrease in Th1 cytokines, GTSF1 participates in disease progression.

5.2.10 Proposed model in CTCL for the role of GTSF1

The points raised in sections **5.2.1** to **5.2.9** have led me to speculate the process in CTCL by which GTSF1 is expressed and its consequences. This model is extrapolated from the references cited above and the experimental data presented here (**Figure 5.1**).

CTCL cells display high heterogeneity that is likely modified by their phenotypic plasticity; phenotypic plasticity is enhanced by the cancer and the memory/effector T cell phenotypes in CTCL. In addition, phenotypic plasticity is modified by epigenetic mechanisms, which are relevant in CTCL. Thus, phenotypic plasticity facilitated by epigenetic mechanisms induces GTSF1 expression. Then, GTSF1 likely forms a complex with tRNAs and other proteins (not shown in model) to perform RNA silencing. This RNA silencing, directly or indirectly, leads to a decreased expression of Th1 associated cytokines IFNγ and TNFα; consequently, the change in cytokine expression drives malignant cells from effector to memory phenotype. This process is likely influenced by the mutational status of the cell, with a particular focus on the NF-κB signaling pathway (not shown in model).

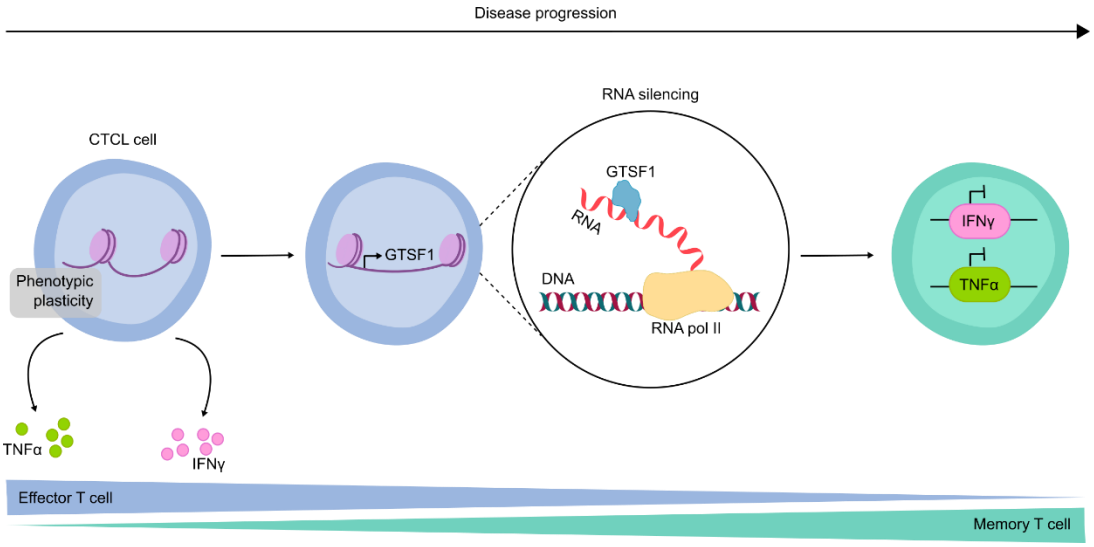


Figure 5.1 Proposed model of GTSF1 expression in CTCL cells

In the early stage of the disease, CTCL cells present a Th1 phenotype with production of IFNγ and TNFα. Phenotypic plasticity then induces GTSF1 expression. GTSF1 forms an RNA silencing complex with tRNAs and other proteins (not shown). The RNA silencing complex is a simplified version of the model presented by Ipsaro and Joshua-Tor [35]. RNA silencing leads, directly or indirectly, to decreased expression of Th1 cytokines. Thus, the initial effector T cell was driven towards a memory T cell phenotype in the late stages of the disease. The figure was created by me in Inkscape with icons available at bioicons.com and Reactome.

5.3 Limitations associated with this study

5.3.1 Biases when planning and performing experimental approaches

As scientists, it is inevitable to bring some cognitive biases to our research [161]. The initial approach in this research project constituted a type of confirmation bias. Confirmation bias is a term used to refer to ways in which our beliefs or expectations influence how we select, retain and evaluate the evidence. In particular, the initial assumption that GTSF1 is participating in TEs control in cancer is a type of confirmation bias called hypothesis-determined information seeking and interpretation [162,163]. In other words, the experimental approach to determine whether GTSF1 performs the same role in cancer as in germ cells constitutes a hypothesis-determined bias. Another type of cognitive bias is functional fixedness. This bias refers what colloquially is known as the inability to "think outside of the box" [164]. In particular, the experimental approach in this project

has been from a cancer biology “box”. However, the research could have been undertaken from different “boxes” such as immunology, developmental biology or dermatology.

In addition, it is important to discuss sex biases. It has been established that immune responses in humans differ between sexes. For example, in females the T cell activation and proliferation is higher, while in males, the number of $CD8^+$ T cells is higher [165]. Considering that CTCL is a malignancy of immune cells, sex can be a determinant for a differential response: It has been reported that CTCL is more commonly diagnosed in males [166]. In this study, only one of the CTCL cell lines used, SZ4, is female-derived; while the lung cancer cell line H1975 is female-derived. In addition, the majority of the CTCL cell lines are male-derived [167]. Therefore, sex is a factor that should be considering when studying CTCL.

5.3.2 Models selected and developed

The following section focuses on my research work in CTCL, however the points mentioned are applicable in the case of lung cancer too. An area of opportunity for this study is the use of other disease models in addition to cell lines. The advantages and disadvantages of cell lines as simplified models of a disease have been discussed thoroughly. Some of the advantages that were considered for the development of my research project are that cell lines are cost effective and easy to use. In addition, as an initial investigation of the molecular mechanism of GTSF1 in cancer, this approach represented a low risk investment. However, the environment in which CTCL cell lines are grown and maintained is not representative of the environment of the disease, i.e. the human body.

In line with the above, the microenvironment in CTCL plays a key role in tumor growth and antitumor immune response. Interactions of CTCL cells with macrophages, dendritic cells,

Langerhans cells, Natural Killer (NK) cells, keratinocytes, endothelial cells, fibroblasts and B cells define the disease course [95,168,169]. Therefore, the interaction of these cells with CTCL cells might modify the results presented here. The inclusion of *in vivo* models represents an interesting area of opportunity for the results presented here.

5.3.3 Context dependency

In light of the results presented here, I hypothesize that the role of GTSF1 can be context dependent. However, the results can also suggest that this phenotype is specific to the cell line Mac2A or pcALCL. For that reason, validation of the GTSF1 knockdown phenotype with other cell lines is of utmost importance. Furthermore, the identification of the factors that determine this phenotype can help evaluate the extrapolation of this phenotype to other contexts.

5.3.4 Current knowledge of GTSF1

GTSF1 was first described in 2007 [36] and the number of publications regarding this protein is variable over the years (**Figure 5.2**). However, the knowledge gap of GTSF1's role under healthy conditions can pose several challenges to understand its role in pathogenesis. The use of high-throughput technologies, unbiased approaches and collaboration across disciplines can help discern its role in carcinogenesis and overcome the challenges this knowledge gap represents.

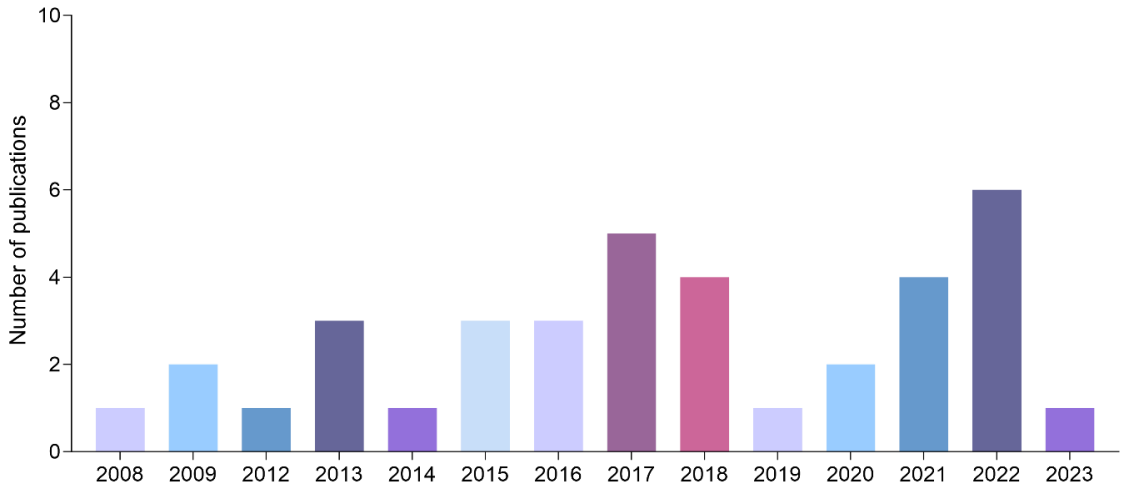


Figure 5.2 Number of publications regarding GTSF1 per year in Pubmed

Pubmed search with the term GTSF1 shows that the number of publications regarding this gene and protein is variable over the years. Pubmed search done on June 6, 2024.

5.4 Proposed future directions

In the Discussion sections above, I briefly introduced some future directions. In the following sections I propose specific experimental approaches to answer the new questions that arise from my thesis work.

5.4.1 Validation of context dependency and refinement of models employed

As mentioned in discussion points **5.3.2** and **5.3.3** the results presented have two major areas of opportunity: to validate the results in additional cell lines and in additional models. The results presented here have the potential to impact the clinical outcome of CTCL patients, however, it is imperative to determine the extent of the phenotype.

As an initial step, validation of the phenotype in additional cell lines can help discern the extent of the phenotype. Specifically, the use of additional cell lines that represent the same variant than Mac2A, pcALCL. Some cell lines are Mac2B, JK and PB2B [167,170-172]. Furthermore, it is important to determine whether this is a phenotype specific for pcALCL or can be extrapolated to

other CTCL variants. An initial approach is to explore whether the phenotype is observed in cell lines that represent other variants of CTCL. Some commonly used cell lines representing SS are Sez4, SeAx, HuT78, H9, Cou-L, Cou-L3, Cou-LS, CTCL-2 and Pno. Some commonly used cell lines representing MF are HH, CTCL-3, MyLa 1929, MyLa 2059, MyLa 2039 and MyLa 3675 [167,173,174].

Currently, the use of *in vivo* models in CTCL remains a challenge. All mice models developed present serious limitations and the majority rely on immunodeficient mice [167]. Considering that GTSF1 knockdown does not affect proliferation nor survival of CTCL cells, xenograft implant of Mac2A SCR and shGTSF1 cells might not provide additional information. In other words, the evidence suggests that tumor growth and volume might not be modified after GTSF1 knockdown. An additional challenge for a xenograft model is that not all CTCL cell lines are able to engraft. In fact, Mac2A has been reported to not form tumors in immunodeficient mice [126]. Therefore, I propose to develop patient derived xenografts from lymph node biopsies or from peripheral blood [175]. Segmentation of patient derived cells with high vs low *GTSF1* expression and engraftment into NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice can help determine whether GTSF1 expression influences acquisition of CTCL clinical manifestations and tumor development. The use of this model would solve the challenge that not all CTCL cell lines are able to engraft but, unfortunately, does not guarantee all patient samples will engraft.

To determine how the microenvironment influences CTCL carcinogenesis, immunocompetent mice models are needed [176]. Therefore, I propose to use the immunocompetent R26STAT3C$^{stopfl/+}$ CD4Cre mouse model [177]. In this model a hyperactive version of STAT3 (termed STAT3C) was knocked into the Rosa26 locus with an upstream floxed stop cassette. Cre recombinase in CD4^{+} cells removes the stop cassette and STAT3C is expressed. This model

presents multiple clinical characteristics that mimic CTCL and its progression such as skin lesions, presence of abnormally looking cells in the blood, clustering of T cells in the skin following a pattern reminiscent of Pautrier microabscesses and, most importantly, inflammatory skin microenvironment. Generating a mouse line with this background and GTSF1 expression would allow to evaluate the role of this gene in CTCL carcinogenesis and progression. One commonly used system is the tetracycline/doxycycline inducible model [178]. Specifically, GTSF1 would be fused to the tetracycline response promoter element (Tet0). In presence of doxycycline the tetracycline-inducible transactivator (rtTA) in $CD4^+$ cells will bind to the Tet0 and lead to GTSF1 expression. Therefore, mice would have a genotype CD4-rtTA/Tet0-GTSF1.

5.4.2 Elucidating the exact molecular mechanism of GTSF1 in carcinogenesis

As discussed in section **5.2.7**, the results in here suggest that GTSF1 participates in the Th1-to-Th2 cytokine shift. This shift has important implications for disease progression and prognosis [71]. Therefore, understanding the exact mechanism in which GTSF1 participates is crucial. Below, I outline some immediate experiments I suggest to answer this question. The results from these experiments will indicate the direction of the next steps.

First, development of GTSF1 knockout CTCL cell line models will eliminate the disadvantages associated with knockdown approaches, particularly the possibility of incomplete knockdown. In line with the work previously done by Yoshimura and colleagues [39] and Ipsaro and colleagues [41], I suggest employing deep sequencing to evaluate changes in sRNAs after GTSF1 knockout. This will enable the evaluation of TEs other than L1, such as Alu A or SVA [10]. Then, to evaluate the consequences of GTSF1 knockout, I suggest performing RNA-Seq, followed by DEGs analysis and pathway enrichment analysis. Taken together, this stronger experimental

approach can give more certainty as to whether GTSF1 participates or not in TEs control and whether it participates in other cellular processes.

Considering that germ cell research suggests GTSF1 binds to RNA [39,41], I suggest identifying its binding partners under malignant conditions. Immunoprecipitation followed by Sequencing (IP-Seq) will allow to identify these binding partners. Additionally, analysis of overlapping genes identified with IP-Seq and RNA-Seq will: (1) validate the results from the analysis suggested in the paragraph above and, (2) signal next directions for understanding the specific role of GTSF1 in CTCL.

5.4.3 GTSF1 as a potential prognosis biomarker

In our lab, *GTSF1* mRNA was initially devised as a possible diagnostic and prognostic biomarker. Considering the recurrent publications of its association with worse prognosis, as well as the results presented here, I suggest further evaluate its use as a prognostic biomarker. An appropriate plan for biomarker validation is key [179]. Some important aspects to consider include the test to be developed (e.g. RT-qPCR or RNA-Seq), the cut-off values for high vs low expression, reproduction in an independent cohort with proper power of the study and a pre-planned statistical analysis.

5.5 Concluding remarks and perspectives

In conclusion, my thesis work over my PhD training has shed light on the role of the germ cell gene GTSF1, in carcinogenesis of CTCL. I have shown that reactivation of developmental programs play a role during carcinogenesis, albeit some might provide passenger (as opposed to driver) phenotypes. In addition, I have shown that GTSF1 is heterogeneously expressed in multiple cancer types. The ectopic GTSF1 expression in lung cancer is potentially associated with TEs

reactivation; the ectopic GTSF1 expression in CTCL is potentially associated with changes in the memory/effector T cell phenotype. I propose that ectopic GTSF1 expression in CTCL cells allows them to hijack a sRNA silencing system, similarly to GTSF1's role in germ cells. In consequence, silencing of target genes leads to changes in the cytokine profile which tilts the phenotype balance from an effector to a memory phenotype. This is in line with the multiple reports of GTSF1 expression and its association with a worse prognosis. Therefore, the use of *GTSF1* mRNA expression levels as a prognostic marker presents a promise for the CTCL field.

References

1. Akers SN, Odunsi K, Karpf AR. Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy. *Future Oncol*. May 2010;6(5):717-32. doi:10.2217/fon.10.36
2. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer*. Aug 2005;5(8):615-25. doi:10.1038/nrc1669
3. van der Bruggen P, Traversari C, Chomez P, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science*. Dec 13 1991;254(5038):1643-7. doi:10.1126/science.1840703
4. Costa FF, Le Blanc K, Brodin B. Concise review: cancer/testis antigens, stem cells, and cancer. *Stem Cells*. Mar 2007;25(3):707-11. doi:10.1634/stemcells.2006-0469
5. Kalejs M, Ivanov A, Plakhins G, et al. Upregulation of meiosis-specific genes in lymphoma cell lines following genotoxic insult and induction of mitotic catastrophe. *BMC Cancer*. Jan 9 2006;6:6. doi:10.1186/1471-2407-6-6
6. Evsikov AV, Marin de Evsikova C. Friend or Foe: Epigenetic Regulation of Retrotransposons in Mammalian Oogenesis and Early Development. *Yale J Biol Med*. Dec 2016;89(4):487-497.
7. Macaulay EC, Chatterjee A, Cheng X, Baguley BC, Eccles MR, Morison IM. The Genes of Life and Death: A Potential Role for Placental-Specific Genes in Cancer: Active retrotransposons in the placenta encode unique functional genes that may also be used by cancer cells to promote malignancy. *Bioessays*. Nov 2017;39(11)doi:10.1002/bies.201700091
8. Ernst C, Odom DT, Kutter C. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nat Commun*. Nov 10 2017;8(1):1411. doi:10.1038/s41467-017-01049-7
9. Ardeljan D, Taylor MS, Ting DT, Burns KH. The Human Long Interspersed Element-1 Retrotransposon: An Emerging Biomarker of Neoplasia. *Clin Chem*. Apr 2017;63(4):816-822. doi:10.1373/clinchem.2016.257444
10. Burns KH. Transposable elements in cancer. *Nat Rev Cancer*. Jul 2017;17(7):415-424. doi:10.1038/nrc.2017.35
11. Nin DS, Deng LW. Biology of Cancer-Testis Antigens and Their Therapeutic Implications in Cancer. *Cells*. Mar 17 2023;12(6)doi:10.3390/cells12060926
12. Fratta E, Coral S, Covre A, et al. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Mol Oncol*. Apr 2011;5(2):164-82. doi:10.1016/j.molonc.2011.02.001
13. Salmaninejad A, Zamani MR, Pourvahedi M, Golchehre Z, Hosseini Bereshneh A, Rezaei N. Cancer/Testis Antigens: Expression, Regulation, Tumor Invasion, and Use in Immunotherapy of Cancers. *Immunol Invest*. Oct 2016;45(7):619-40. doi:10.1080/08820139.2016.1197241
14. Anwar SL, Wulaningsih W, Lehmann U. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *Int J Mol Sci*. May 4 2017;18(5)doi:10.3390/ijms18050974
15. Singh A, Rappolee DA, Ruden DM. Epigenetic Reprogramming in Mice and Humans: From Fertilization to Primordial Germ Cell Development. *Cells*. Jul 17 2023;12(14)doi:10.3390/cells12141874

16. Miller JL, Grant PA. The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcell Biochem*. 2013;61:289-317. doi:10.1007/978-94-007-4525-4_13
17. Odroniec A, Olszewska M, Kurpisz M. Epigenetic markers in the embryonal germ cell development and spermatogenesis. *Basic Clin Androl*. Feb 23 2023;33(1):6. doi:10.1186/s12610-022-00179-3
18. Bozdemir N, Uysal F. Histone acetyltransferases and histone deacetyl transferases play crucial role during oogenesis and early embryo development. *Genesis*. Sep 2023;61(5):e23518. doi:10.1002/dvg.23518
19. Kalejs M, Erenpreisa J. Cancer/testis antigens and gametogenesis: a review and "brain-storming" session. *Cancer Cell Int*. Feb 16 2005;5(1):4. doi:10.1186/1475-2867-5-4
20. Fardi M, Solali S, Farshdousti Hagh M. Epigenetic mechanisms as a new approach in cancer treatment: An updated review. *Genes Dis*. Dec 2018;5(4):304-311. doi:10.1016/j.gendis.2018.06.003
21. Wilkinson AL, Zorzan I, Rugg-Gunn PJ. Epigenetic regulation of early human embryo development. *Cell Stem Cell*. 2023;30(12):1569-1584. doi:10.1016/j.stem.2023.09.010
22. Gantchev J, Messina-Pacheco J, Martinez Villarreal A, et al. Ectopically Expressed Meiosis-Specific Cancer Testis Antigen HORMAD1 Promotes Genomic Instability in Squamous Cell Carcinomas. *Cells*. Jun 14 2023;12(12)doi:10.3390/cells12121627
23. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. Mar 4 2011;144(5):646-74. doi:10.1016/j.cell.2011.02.013
24. Martin SL. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J Biomed Biotechnol*. 2006;2006(1):45621. doi:10.1155/JBB/2006/45621
25. Serrato-Capuchina A, Matute DR. The Role of Transposable Elements in Speciation. *Genes (Basel)*. May 15 2018;9(5)doi:10.3390/genes9050254
26. Cajuso T, Sulo P, Tanskanen T, et al. Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nat Commun*. Sep 6 2019;10(1):4022. doi:10.1038/s41467-019-11770-0
27. Doucet AJ, Hulme AE, Sahinovic E, et al. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet*. Oct 7 2010;6(10)doi:10.1371/journal.pgen.1001150
28. Forrester JV, Xu H, Lambe T, Cornall R. Immune privilege or privileged immunity? *Mucosal Immunol*. Sep 2008;1(5):372-81. doi:10.1038/mi.2008.27
29. Gure AO, Stockert E, Arden KC, et al. CT10: a new cancer-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational-difference analysis. *Int J Cancer*. Mar 1 2000;85(5):726-32. doi:10.1002/(sici)1097-0215(20000301)85:5<726::aid-ijc21>3.0.co;2-f
30. Liang Y, Qu X, Shah NM, Wang T. Towards targeting transposable elements for cancer therapy. *Nat Rev Cancer*. Feb 2024;24(2):123-140. doi:10.1038/s41568-023-00653-8
31. Gjerstorff MF, Andersen MH, Ditzel HJ. Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget*. Jun 30 2015;6(18):15772-87. doi:10.18632/oncotarget.4694
32. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*. Feb 2019;20(2):89-108. doi:10.1038/s41576-018-0073-3
33. Han YN, Li Y, Xia SQ, Zhang YY, Zheng JH, Li W. PIWI Proteins and PIWI-Interacting RNA: Emerging Roles in Cancer. *Cell Physiol Biochem*. 2017;44(1):1-20. doi:10.1159/000484541

34. Chetta M, Di Pietro L, Bukvic N, Lattanzi W. Rising Roles of Small Noncoding RNAs in Cotranscriptional Regulation: In Silico Study of miRNA and piRNA Regulatory Network in Humans. *Genes (Basel)*. Apr 29 2020;11(5)doi:10.3390/genes11050482
35. Ipsaro JJ, Joshua-Tor L. Developmental roles and molecular mechanisms of Asterix/GTSF1. *Wiley Interdiscip Rev RNA*. Sep 2022;13(5):e1716. doi:10.1002/wrna.1716
36. Yoshimura T, Miyazaki T, Toyoda S, et al. Gene expression pattern of Cue110: a member of the uncharacterized UPF0224 gene family preferentially expressed in germ cells. *Gene Expr Patterns*. Dec 2007;8(1):27-35. doi:10.1016/j.modgep.2007.08.003
37. Huntriss J, Lu J, Hemmings K, et al. Isolation and expression of the human gametocyte-specific factor 1 gene (GTSF1) in fetal ovary, oocytes, and preimplantation embryos. *J Assist Reprod Genet*. Jan 2017;34(1):23-31. doi:10.1007/s10815-016-0795-0
38. Hadziselimovic F, Hadziselimovic NO, Demougin P, Krey G, Oakeley E. Piwi-pathway alteration induces LINE-1 transposon derepression and infertility development in cryptorchidism. *Sex Dev*. 2015;9(2):98-104. doi:10.1159/000375351
39. Yoshimura T, Watanabe T, Kuramochi-Miyagawa S, et al. Mouse GTSF1 is an essential factor for secondary piRNA biogenesis. *EMBO Rep*. Apr 2018;19(4)doi:10.15252/embr.201642054
40. Yoshimura T, Toyoda S, Kuramochi-Miyagawa S, et al. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev Biol*. Nov 1 2009;335(1):216-27. doi:10.1016/j.ydbio.2009.09.003
41. Ipsaro JJ, O'Brien PA, Bhattacharya S, Palmer AG, 3rd, Joshua-Tor L. Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons. *Cell Rep*. Mar 30 2021;34(13):108914. doi:10.1016/j.celrep.2021.108914
42. Litvinov IV, Cordeiro B, Huang Y, et al. Ectopic expression of cancer-testis antigens in cutaneous T-cell lymphoma patients. *Clinical cancer research : an official journal of the American Association for Cancer Research*. Jul 15 2014;20(14):3799-808. doi:10.1158/1078-0432.Ccr-14-0307
43. Litvinov IV, Netchiporouk E, Cordeiro B, et al. The Use of Transcriptional Profiling to Improve Personalized Diagnosis and Management of Cutaneous T-cell Lymphoma (CTCL). *Clinical cancer research : an official journal of the American Association for Cancer Research*. Jun 15 2015;21(12):2820-9. doi:10.1158/1078-0432.CCR-14-3322
44. Litvinov IV, Tetzlaff MT, Thibault P, et al. Gene expression analysis in Cutaneous T-Cell Lymphomas (CTCL) highlights disease heterogeneity and potential diagnostic and prognostic indicators. *Oncoimmunology*. 2017;6(5):e1306618. doi:10.1080/2162402X.2017.1306618
45. Liu X, Jin S, Hu S, et al. Single-cell transcriptomics links malignant T cells to the tumor immune landscape in cutaneous T cell lymphoma. *Nat Commun*. Mar 3 2022;13(1):1158. doi:10.1038/s41467-022-28799-3
46. Rindler K, Jonak C, Alkon N, et al. Single-cell RNA sequencing reveals markers of disease progression in primary cutaneous T-cell lymphoma. *Mol Cancer*. Sep 28 2021;20(1):124. doi:10.1186/s12943-021-01419-2
47. van Kester MS, Borg MK, Zoutman WH, et al. A meta-analysis of gene expression data identifies a molecular signature characteristic for tumor-stage mycosis fungoides. *The Journal of investigative dermatology*. Aug 2012;132(8):2050-9. doi:10.1038/jid.2012.117
48. Xiao MZX, Hennessey D, Iyer A, et al. Transcriptomic changes during stage progression of mycosis fungoides. *Br J Dermatol*. Mar 2022;186(3):520-531. doi:10.1111/bjd.20760

49. Lefrancois P, Tetzlaff MT, Moreau L, et al. TruSeq-Based Gene Expression Analysis of Formalin-Fixed Paraffin-Embedded (FFPE) Cutaneous T-Cell Lymphoma Samples: Subgroup Analysis Results and Elucidation of Biases from FFPE Sample Processing on the TruSeq Platform. *Frontiers in medicine*. 2017;4:153. doi:10.3389/fmed.2017.00153
50. Becker H, Marcucci G, Maharry K, et al. Mutations of the Wilms tumor 1 gene (WT1) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood*. Aug 5 2010;116(5):788-92. doi:10.1182/blood-2010-01-262543
51. Zhang C, Bai G, Zhu W, Bai D, Bi G. Identification of miRNA-mRNA Network Associated with Acute Myeloid Leukemia Survival. *Med Sci Monit*. Oct 1 2017;23:4705-4714. doi:10.12659/msm.903989
52. Gao DY, Ling Y, Lou XL, Wang YY, Liu LM. GTSF1 gene may serve as a novel potential diagnostic biomarker for liver cancer. *Oncol Lett*. Mar 2018;15(3):3133-3140. doi:10.3892/ol.2017.7695
53. Simoes MCF, Sousa JJS, Pais A. Skin cancer and new treatment perspectives: a review. *Cancer Lett*. Feb 1 2015;357(1):8-42. doi:10.1016/j.canlet.2014.11.001
54. Bagherani N, Smoller BR. An overview of cutaneous T cell lymphomas. *F1000Res*. 2016;5doi:10.12688/f1000research.8829.1
55. Bradford PT, Devesa SS, Anderson WF, Toro JR. Cutaneous lymphoma incidence patterns in the United States: a population-based study of 3884 cases. *Blood*. May 21 2009;113(21):5064-73. doi:10.1182/blood-2008-10-184168
56. Jawed SI, Myskowski PL, Horwitz S, Moskowitz A, Querfeld C. Primary cutaneous T-cell lymphoma (mycosis fungoides and Sezary syndrome): part I. Diagnosis: clinical and histopathologic features and new molecular and biologic markers. *J Am Acad Dermatol*. Feb 2014;70(2):205 e1-16; quiz 221-2. doi:10.1016/j.jaad.2013.07.049
57. Wilson LD, Hinds GA, Yu JB. Age, race, sex, stage, and incidence of cutaneous lymphoma. *Clin Lymphoma Myeloma Leuk*. Oct 2012;12(5):291-6. doi:10.1016/j.clml.2012.06.010
58. Torrealba MP, Manfrere KC, Miyashiro DR, et al. Chronic activation profile of circulating CD8+ T cells in Sezary syndrome. *Oncotarget*. Jan 9 2018;9(3):3497-3506. doi:10.18632/oncotarget.23334
59. Litvinov IV, Shtreis A, Kobayashi K, et al. Investigating potential exogenous tumor initiating and promoting factors for Cutaneous T-Cell Lymphomas (CTCL), a rare skin malignancy. *Oncoimmunology*. Jul 2016;5(7):e1175799. doi:10.1080/2162402X.2016.1175799
60. Olsen E, Vonderheid E, Pimpinelli N, et al. Revisions to the staging and classification of mycosis fungoides and Sezary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC). *Blood*. Sep 15 2007;110(6):1713-22. doi:10.1182/blood-2007-03-055749
61. Girardi M, Heald PW, Wilson LD. The pathogenesis of mycosis fungoides. *N Engl J Med*. May 6 2004;350(19):1978-88. doi:10.1056/NEJMra032810
62. Martinez Villarreal A, Gantchev J, Lagace F, et al. Hypopigmented Mycosis Fungoides: Loss of Pigmentation Reflects Antitumor Immune Response in Young Patients. *Cancers (Basel)*. Jul 22 2020;12(8)doi:10.3390/cancers12082007
63. Li JY, Horwitz S, Moskowitz A, Myskowski PL, Pulitzer M, Querfeld C. Management of cutaneous T cell lymphoma: new and emerging targets and treatment options. *Cancer Manag Res*. 2012;4:75-89. doi:10.2147/CMAR.S9660

64. Pina-Oviedo S, Ortiz-Hidalgo C, Carballo-Zarate AA, Zarate-Osorno A. ALK-Negative Anaplastic Large Cell Lymphoma: Current Concepts and Molecular Pathogenesis of a Heterogeneous Group of Large T-Cell Lymphomas. *Cancers (Basel)*. Sep 17 2021;13(18)doi:10.3390/cancers13184667
65. Querfeld C, Zain J, Rosen ST. Primary Cutaneous T-Cell Lymphomas: Mycosis Fungoides and Sezary Syndrome. *Cancer Treat Res*. 2019;176:225-248. doi:10.1007/978-3-319-99716-2_11
66. Damsky WE, Choi J. Genetics of Cutaneous T Cell Lymphoma: From Bench to Bedside. *Curr Treat Options Oncol*. Jul 2016;17(7):33. doi:10.1007/s11864-016-0410-8
67. Olsen EA, Whittaker S, Willemze R, et al. Primary cutaneous lymphoma: recommendations for clinical trial design and staging update from the ISCL, USCLC, and EORTC. *Blood*. 2022;140(5):419-437. doi:10.1182/blood.2021012057
68. Zhang XR, Chien PN, Nam SY, Heo CY. Anaplastic Large Cell Lymphoma: Molecular Pathogenesis and Treatment. *Cancers (Basel)*. Mar 24 2022;14(7)doi:10.3390/cancers14071650
69. Wilcox RA. Cutaneous T-cell lymphoma: 2017 update on diagnosis, risk-stratification, and management. *Am J Hematol*. Oct 2017;92(10):1085-1102. doi:10.1002/ajh.24876
70. Jawed SI, Myskowski PL, Horwitz S, Moskowitz A, Querfeld C. Primary cutaneous T-cell lymphoma (mycosis fungoides and Sezary syndrome): part II. Prognosis, management, and future directions. *J Am Acad Dermatol*. Feb 2014;70(2):223 e1-17; quiz 240-2. doi:10.1016/j.jaad.2013.08.033
71. Kim EJ, Hess S, Richardson SK, et al. Immunopathogenesis and therapy of cutaneous T cell lymphoma. *J Clin Invest*. Apr 2005;115(4):798-812. doi:10.1172/JCI24826
72. Liszewski W, Naym DG, Biskup E, Gniadecki R. Psoralen with ultraviolet A-induced apoptosis of cutaneous lymphoma cell lines is augmented by type I interferons via the JAK1-STAT1 pathway. *Photodermatol Photoimmunol Photomed*. May 2017;33(3):164-171. doi:10.1111/phpp.12302
73. Elenitoba-Johnson KS, Wilcox R. A new molecular paradigm in mycosis fungoides and Sezary syndrome. *Semin Diagn Pathol*. Jan 2017;34(1):15-21. doi:10.1053/j.semdp.2016.11.002
74. Phyo ZH, Shanbhag S, Rozati S. Update on Biology of Cutaneous T-Cell Lymphoma. *Front Oncol*. 2020;10:765. doi:10.3389/fonc.2020.00765
75. Gluud M, Willerslev-Olsen A, Gjerdrum LMR, et al. MicroRNAs in the Pathogenesis, Diagnosis, Prognosis and Targeted Treatment of Cutaneous T-Cell Lymphomas. *Cancers (Basel)*. May 13 2020;12(5)doi:10.3390/cancers12051229
76. Izykowska K. Methylation patterns of cutaneous T-cell lymphomas. *Experimental dermatology*. Apr 30 2020;doi:10.1111/exd.14108
77. Litvinov IV, Cordeiro B, Fredholm S, et al. Analysis of STAT4 expression in cutaneous T-cell lymphoma (CTCL) patients and patient-derived cell lines. *Cell cycle (Georgetown, Tex)*. 2014;13(18):2975-82. doi:10.4161/15384101.2014.947759
78. Huang Y, Litvinov IV, Wang Y, et al. Thymocyte selection-associated high mobility group box gene (TOX) is aberrantly over-expressed in mycosis fungoides and correlates with poor prognosis. *Oncotarget*. Jun 30 2014;5(12):4418-25. doi:10.18632/oncotarget.2031
79. Dulmage BO, Akilov O, Vu JR, Falo LD, Geskin LJ. Dysregulation of the TOX-RUNX3 pathway in cutaneous T-cell lymphoma. *Oncotarget*. May 3 2019;10(33):3104-3113. doi:10.18632/oncotarget.5742
80. Litvinov IV, Netchiporouk E, Cordeiro B, et al. Ectopic expression of embryonic stem cell and other developmental genes in cutaneous T-cell lymphoma. *Oncoimmunology*. Nov 2014;3(11):e970025. doi:10.4161/21624011.2014.970025

81. Iyer A, Hennessey D, O'Keefe S, et al. Branched evolution and genomic intratumor heterogeneity in the pathogenesis of cutaneous T-cell lymphoma. *Blood Adv*. Jun 9 2020;4(11):2489-2500. doi:10.1182/bloodadvances.2020001441
82. Iyer A, Hennessey D, O'Keefe S, et al. Skin colonization by circulating neoplastic clones in cutaneous T-cell lymphoma. *Blood*. Oct 31 2019;134(18):1517-1527. doi:10.1182/blood.2019002516
83. Iyer A, Hennessey D, O'Keefe S, et al. Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Adv*. Apr 9 2019;3(7):1175-1184. doi:10.1182/bloodadvances.2018027482
84. Harro CM, Sprenger KB, Chaurio RA, et al. Sezary syndrome originates from heavily mutated hematopoietic progenitors. *Blood Adv*. Aug 2 2023;doi:10.1182/bloodadvances.2022008562
85. Kunzli M, Masopust D. CD4(+) T cell memory. *Nat Immunol*. Jun 2023;24(6):903-914. doi:10.1038/s41590-023-01510-4
86. Campbell JJ, Clark RA, Watanabe R, Kupper TS. Sezary syndrome and mycosis fungoides arise from distinct T-cell subsets: a biologic rationale for their distinct clinical behaviors. *Blood*. Aug 5 2010;116(5):767-71. doi:10.1182/blood-2009-11-251926
87. Khalil S, Bardawil T, Kurban M, Abbas O. Tissue-resident memory T cells in the skin. *Inflamm Res*. Mar 2020;69(3):245-254. doi:10.1007/s00011-020-01320-6
88. Tokura Y, Phadungsaksawasdi P, Kurihara K, Fujiyama T, Honda T. Pathophysiology of Skin Resident Memory T Cells. *Front Immunol*. 2020;11:618897. doi:10.3389/fimmu.2020.618897
89. Kaech SM, Cui W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol*. Nov 2012;12(11):749-61. doi:10.1038/nri3307
90. Fonseca R, Beura LK, Quarnstrom CF, et al. Developmental plasticity allows outside-in immune responses by resident memory T cells. *Nat Immunol*. Apr 2020;21(4):412-421. doi:10.1038/s41590-020-0607-7
91. Poglio S, Prochazkova-Carlotti M, Cherrier F, et al. Xenograft and cell culture models of Sezary syndrome reveal cell of origin diversity and subclonal heterogeneity. *Leukemia*. Jun 2021;35(6):1696-1709. doi:10.1038/s41375-020-01068-2
92. Guenova E, Watanabe R, Teague JE, et al. TH2 cytokines from malignant cells suppress TH1 responses and enforce a global TH2 bias in leukemic cutaneous T-cell lymphoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. Jul 15 2013;19(14):3755-63. doi:10.1158/1078-0432.CCR-12-3488
93. Nakajima R, Miyagaki T, Hirakawa M, et al. Interleukin-25 is involved in cutaneous T-cell lymphoma progression by establishing a T helper 2-dominant microenvironment. *Br J Dermatol*. Jun 2018;178(6):1373-1382. doi:10.1111/bjd.16237
94. Jankowska-Konsur A, Kobierzycki C, Grzegrzolka J, et al. Podoplanin Expression Correlates with Disease Progression in Mycosis Fungoides. *Acta Derm Venereol*. Feb 8 2017;97(2):235-241. doi:10.2340/00015555-2517
95. Dobos G, Lazaridou I, de Masson A. Mycosis Fungoides and Sezary Syndrome: Microenvironment and Cancer Progression. *Cancers (Basel)*. Jan 25 2023;15(3)doi:10.3390/cancers15030746
96. Rowden G, Phillips TM, Lewis MG, Wilkinson RD. Target role of Langerhans cells in mycosis fungoides: transmission and immuno-electron microscopic studies. *J Cutan Pathol*. Oct 1979;6(5):364-82. doi:10.1111/j.1600-0560.1979.tb01159.x

97. Furlan FC, Sanches JA. Hypopigmented mycosis fungoides: a review of its clinical features and pathophysiology. *An Bras Dermatol*. Nov-Dec 2013;88(6):954-60. doi:10.1590/abd1806-4841.20132336
98. Ardigo M, Borroni G, Muscardin L, Kerl H, Cerroni L. Hypopigmented mycosis fungoides in Caucasian patients: a clinicopathologic study of 7 cases. *J Am Acad Dermatol*. Aug 2003;49(2):264-70.
99. Howard MS, Smoller BR. Mycosis fungoides: Classic disease and variant presentations. *Semin Cutan Med Surg*. Jun 2000;19(2):91-99. doi:Doi 10.1016/S1085-5629(00)80005-X
100. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*. 04/05 2003;124
101. Liu F, Gao Y, Xu B, et al. PEG10 amplification at 7q21.3 potentiates large-cell transformation in cutaneous T-cell lymphoma. *Blood*. Jan 27 2022;139(4):554-571. doi:10.1182/blood.2021012091
102. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. Jan 1 2002;30(1):207-10. doi:10.1093/nar/30.1.207
103. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*. 2012;41(D1):D991-D995. doi:10.1093/nar/gks1193
104. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. Aug 1 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170
105. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*. 2018;7:1338. doi:10.12688/f1000research.15931.2
106. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. May 2016;34(5):525-7. doi:10.1038/nbt.3519
107. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. Sep 1 2011;27(17):2325-9. doi:10.1093/bioinformatics/btr355
108. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521. doi:10.12688/f1000research.7563.2
109. Wang Z, Liu N, Shi S, Liu S, Lin H. The Role of PIWIL4, an Argonaute Family Protein, in Breast Cancer. *J Biol Chem*. May 13 2016;291(20):10646-58. doi:10.1074/jbc.M116.723239
110. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH, Jr. High frequency retrotransposition in cultured mammalian cells. *Cell*. Nov 29 1996;87(5):917-27. doi:10.1016/s0092-8674(00)81998-4
111. Xie Y, Rosser JM, Thompson TL, Boeke JD, An W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res*. Feb 2011;39(3):e16. doi:10.1093/nar/gkq1076
112. Tsang M, Gantchev J, Ghazawi FM, Litvinov IV. Protocol for adhesion and immunostaining of lymphocytes and other non-adherent cells in culture. *Biotechniques*. Nov 1 2017;63(5):230-233. doi:10.2144/000114610
113. Gantchev J, Ramchatesingh B, Berman-Rosa M, et al. Tools used to assay genomic instability in cancers and cancer meiomitosis. *J Cell Commun Signal*. Jun 2022;16(2):159-177. doi:10.1007/s12079-021-00661-z

114. Mellacheruvu D, Wright Z, Couzens AL, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*. Aug 2013;10(8):730-6. doi:10.1038/nmeth.2557
115. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. Jul 2007;35(Web Server issue):W193-200. doi:10.1093/nar/gkm226
116. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res*. Jul 5 2023;51(W1):W207-W212. doi:10.1093/nar/gkad347
117. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. Jan 6 2023;51(D1):D638-D646. doi:10.1093/nar/gkac1000
118. Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. Feb 2019;14(2):482-517. doi:10.1038/s41596-018-0103-9
119. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. Nov 2019;28(11):1947-1951. doi:10.1002/pro.3715
120. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. Jan 6 2023;51(D1):D587-D592. doi:10.1093/nar/gkac963
121. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. Jan 1 2000;28(1):27-30. doi:10.1093/nar/28.1.27
122. Tsang M, Gantchev J, Netchiporouk E, et al. A study of meiomitosis and novel pathways of genomic instability in cutaneous T-cell lymphomas (CTCL). *Oncotarget*. Dec 28 2018;9(102):37647-37661. doi:10.18632/oncotarget.26479
123. Touma F, Lambert M, Martinez Villarreal A, Gantchev J, Ramchatesingh B, Litvinov IV. The Ultraviolet Irradiation of Keratinocytes Induces Ectopic Expression of LINE-1 Retrotransposon Machinery and Leads to Cellular Senescence. *Biomedicines*. Nov 10 2023;11(11)doi:10.3390/biomedicines11113017
124. Genzor P, Cordts SC, Bokil NV, Haase AD. Aberrant expression of select piRNA-pathway genes does not reactivate piRNA silencing in cancer cells. *Proc Natl Acad Sci U S A*. Jun 4 2019;116(23):11111-11112. doi:10.1073/pnas.1904498116
125. Bodak M, Yu J, Ciaudo C. Regulation of LINE-1 in mammals. *Biomol Concepts*. Oct 2014;5(5):409-28. doi:10.1515/bmc-2014-0018
126. Netchiporouk E, Gantchev J, Tsang M, et al. Analysis of CTCL cell lines reveals important differences between mycosis fungoides/Sezary syndrome vs. HTLV-1(+) leukemic cell lines. *Oncotarget*. Nov 10 2017;8(56):95981-95998. doi:10.18632/oncotarget.21619
127. Minocherhomji S, Liu Y, He YD, Fielden MR. Biomarkers of genome instability in normal mammalian genomes following drug-induced replication stress. *Environ Mol Mutagen*. Oct 2020;61(8):770-785. doi:10.1002/em.22364
128. Zhao X, Zhang X, Shao S, et al. High expression of GMNN predicts malignant progression and poor prognosis in ACC. *Eur J Med Res*. Dec 20 2022;27(1):301. doi:10.1186/s40001-022-00950-2
129. Migliazza A, Lombardi L, Rocchi M, et al. Heterogeneous chromosomal aberrations generate 3' truncations of the NFKB2/lyt-10 gene in lymphoid malignancies. *Blood*. Dec 1 1994;84(11):3850-60.

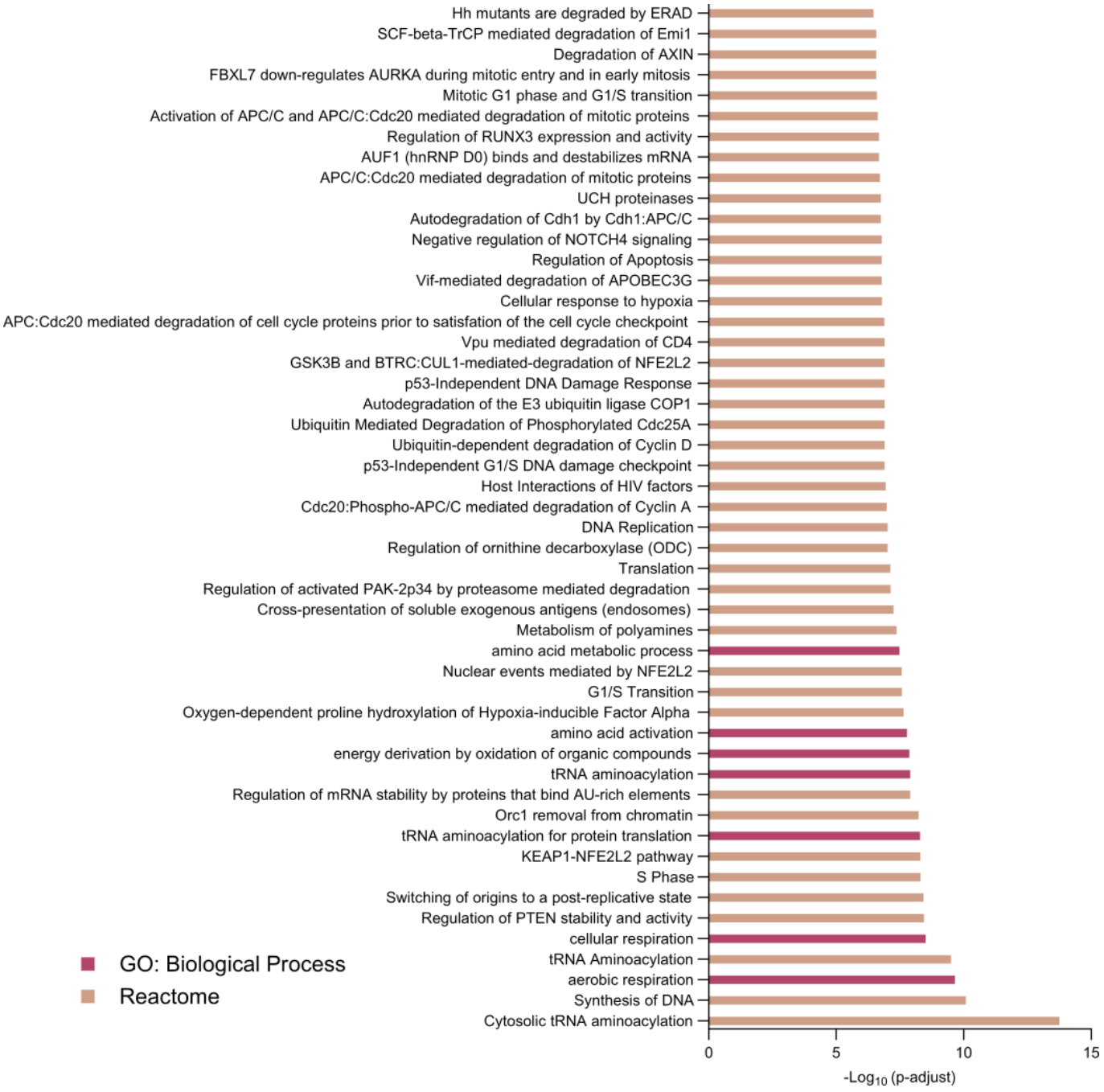
130. Zhang J, Chang CC, Lombardi L, Dalla-Favera R. Rearranged NFKB2 gene in the HUT78 T-lymphoma cell line codes for a constitutively nuclear factor lacking transcriptional repressor functions. *Oncogene*. Jul 1994;9(7):1931-7.
131. Reina-Campos M, Scharping NE, Goldrath AW. CD8(+) T cell metabolism in infection and cancer. *Nat Rev Immunol*. Nov 2021;21(11):718-738. doi:10.1038/s41577-021-00537-8
132. Mehta AK, Gracias DT, Croft M. TNF activity and T cells. *Cytokine*. Jan 2018;101:14-18. doi:10.1016/j.cyto.2016.08.003
133. Blanchett S, Boal-Carvalho I, Layzell S, Seddon B. NF-kappaB and Extrinsic Cell Death Pathways - Entwined Do-or-Die Decisions for T cells. *Trends Immunol*. Jan 2021;42(1):76-88. doi:10.1016/j.it.2020.10.013
134. Zhao Y, Shao Q, Peng G. Exhaustion and senescence: two crucial dysfunctional states of T cells in the tumor microenvironment. *Cell Mol Immunol*. Jan 2020;17(1):27-35. doi:10.1038/s41423-019-0344-8
135. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. Jan 2022;12(1):31-46. doi:10.1158/2159-8290.CD-21-1059
136. Clayton EA, Wang L, Rishishwar L, Wang J, McDonald JF, Jordan IK. Patterns of Transposable Element Expression and Insertion in Cancer. *Front Mol Biosci*. 2016;3:76. doi:10.3389/fmolb.2016.00076
137. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood*. Sep 15 2022;140(11):1229-1253. doi:10.1182/blood.2022015851
138. Lefrancois P, Xie P, Wang L, et al. Gene expression profiling and immune cell-type deconvolution highlight robust disease progression and survival markers in multiple cohorts of CTCL patients. *Oncoimmunology*. 2018;7(8):e1467856. doi:10.1080/2162402X.2018.1467856
139. Litvinov IV, Jones DA, Sasseville D, Kupper TS. Transcriptional profiles predict disease outcome in patients with cutaneous T-cell lymphoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. Apr 1 2010;16(7):2106-14. doi:10.1158/1078-0432.Ccr-09-2879
140. Garcia-Borja E, Siegl F, Mateu R, et al. Critical appraisal of the piRNA-PIWI axis in cancer and cancer stem cells. *Biomark Res*. Feb 1 2024;12(1):15. doi:10.1186/s40364-024-00563-3
141. Lu Y, Zhang K, Li C, et al. Piwil2 suppresses p53 by inducing phosphorylation of signal transducer and activator of transcription 3 in tumor cells. *PLoS One*. 2012;7(1):e30999. doi:10.1371/journal.pone.0030999
142. Tajeddine N, Gala JL, Louis M, Van Schoor M, Tombal B, Gailly P. Tumor-associated antigen preferentially expressed antigen of melanoma (PRAME) induces caspase-independent cell death in vitro and reduces tumorigenicity in vivo. *Cancer Res*. Aug 15 2005;65(16):7348-55. doi:10.1158/0008-5472.CAN-04-4011
143. Daskalos A, Nikolaidis G, Xinarianos G, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer*. Jan 1 2009;124(1):81-7. doi:10.1002/ijc.23849
144. Andor N, Maley CC, Ji HP. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Res*. May 1 2017;77(9):2179-2185. doi:10.1158/0008-5472.CAN-16-1553
145. Bamezai S, Pulikkottil AJ, Yadav T, et al. A noncanonical enzymatic function of PIWIL4 maintains genomic integrity and leukemic growth in AML. *Blood*. Jul 6 2023;142(1):90-105. doi:10.1182/blood.2022018651

146. Li Z, Li Z, Zhang Y, et al. Mammalian PIWI-piRNA-target complexes reveal features for broad and efficient target silencing. *Nat Struct Mol Biol*. Apr 24 2024;doi:10.1038/s41594-024-01287-6
147. Buus TB, Willerslev-Olsen A, Fredholm S, et al. Single-cell heterogeneity in Sezary syndrome. *Blood Adv*. Aug 28 2018;2(16):2115-2126. doi:10.1182/bloodadvances.2018022608
148. Horna P, Moscinski LC, Sokol L, Shao H. Naive/memory T-cell phenotypes in leukemic cutaneous T-cell lymphoma: Putative cell of origin overlaps disease classification. *Cytometry B Clin Cytom*. May 2019;96(3):234-241. doi:10.1002/cyto.b.21738
149. Iyer A, Hennessey D, O'Keefe S, et al. Independent evolution of cutaneous lymphoma subclones in different microenvironments of the skin. *Sci Rep*. Sep 23 2020;10(1):15483. doi:10.1038/s41598-020-72459-9
150. Strobl J, Haniffa M. Functional heterogeneity of human skin-resident memory T cells in health and disease. *Immunol Rev*. May 5 2023;doi:10.1111/imr.13213
151. Watanabe M, Sasaki M, Itoh K, et al. JunB induced by constitutive CD30-extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase signaling activates the CD30 promoter in anaplastic large cell lymphoma and reed-sternberg cells of Hodgkin lymphoma. *Cancer Res*. Sep 1 2005;65(17):7628-34. doi:10.1158/0008-5472.CAN-05-0925
152. Gerondakis S, Fulford TS, Messina NL, Grumont RJ. NF-kappaB control of T cell development. *Nat Immunol*. Jan 2014;15(1):15-25. doi:10.1038/ni.2785
153. Daniels MA, Luera D, Teixeiro E. NFkappaB signaling in T cell memory. *Front Immunol*. 2023;14:1129191. doi:10.3389/fimmu.2023.1129191
154. Ungewickell A, Bhaduri A, Rios E, et al. Genomic analysis of mycosis fungoides and Sezary syndrome identifies recurrent alterations in TNFR2. *Nat Genet*. Sep 2015;47(9):1056-60. doi:10.1038/ng.3370
155. da Silva Almeida AC, Abate F, Khiabanian H, et al. The mutational landscape of cutaneous T cell lymphoma and Sezary syndrome. *Nat Genet*. Dec 2015;47(12):1465-70. doi:10.1038/ng.3442
156. Wang L, Ni X, Covington KR, et al. Genomic profiling of Sezary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat Genet*. Dec 2015;47(12):1426-34. doi:10.1038/ng.3444
157. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet*. Sep 2015;47(9):1011-9. doi:10.1038/ng.3356
158. Wang H, Wei W, Zhang JP, et al. A novel model of alternative NF-kappaB pathway activation in anaplastic large cell lymphoma. *Leukemia*. Jul 2021;35(7):1976-1989. doi:10.1038/s41375-020-01088-y
159. Ren J, Qu R, Rahman NT, et al. Integrated transcriptome and trajectory analysis of cutaneous T-cell lymphoma identifies putative precancer populations. *Blood Adv*. Feb 14 2023;7(3):445-457. doi:10.1182/bloodadvances.2022008168
160. Ai W, Li H, Song N, Li L, Chen H. Optimal method to stimulate cytokine production and its use in immunotoxicity assessment. *Int J Environ Res Public Health*. Aug 27 2013;10(9):3834-42. doi:10.3390/ijerph10093834
161. Hudson R. Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis. *J Gen Philos Sci*. 2021;52(3):389-405. doi:10.1007/s10838-020-09548-w
162. Peters U. What Is the Function of Confirmation Bias? *Erkenntnis*. 2022/06/01 2022;87(3):1351-1376. doi:10.1007/s10670-020-00252-1

163. Hahn U, Harris AJL. Chapter Two - What Does It Mean to be Biased: Motivated Reasoning and Rationality. In: Ross BH, ed. *Psychology of Learning and Motivation*. Academic Press; 2014:41-102.
164. Chrysikou EG, Motyka K, Nigro C, Yang SI, Thompson-Schill SL. Functional Fixedness in Creative Thinking Tasks Depends on Stimulus Modality. *Psychol Aesthet Creat Arts*. Nov 2016;10(4):425-435. doi:10.1037/aca0000050
165. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. Oct 2016;16(10):626-38. doi:10.1038/nri.2016.90
166. Ghazawi FM, Alghazawi N, Le M, et al. Environmental and Other Extrinsic Risk Factors Contributing to the Pathogenesis of Cutaneous T Cell Lymphoma (CTCL). *Front Oncol*. 2019;9:300. doi:10.3389/fonc.2019.00300
167. Gill RPK, Gantchev J, Martinez Villarreal A, et al. Understanding Cell Lines, Patient-Derived Xenograft and Genetically Engineered Mouse Models Used to Study Cutaneous T-Cell Lymphoma. *Cells*. Feb 9 2022;11(4)doi:10.3390/cells11040593
168. Gaydosik AM, Stonesifer CJ, Tabib T, Lafyatis R, Geskin LJ, Fuschiotti P. The mycosis fungoides cutaneous microenvironment shapes dysfunctional cell trafficking, antitumor immunity, matrix interactions, and angiogenesis. *JCI Insight*. Oct 9 2023;8(19)doi:10.1172/jci.insight.170015
169. Rubio Gonzalez B, Zain J, Rosen ST, Querfeld C. Tumor microenvironment in mycosis fungoides and Sezary syndrome. *Curr Opin Oncol*. Jan 2016;28(1):88-96. doi:10.1097/CCO.0000000000000243
170. Drexler HG, MacLeod RA. Malignant hematopoietic cell lines: in vitro models for the study of anaplastic large-cell lymphoma. *Leukemia*. Oct 2004;18(10):1569-71. doi:10.1038/sj.leu.2403465
171. Schiemann WP, Pfeifer WM, Levi E, Kadin ME, Lodish HF. A deletion in the gene for transforming growth factor beta type I receptor abolishes growth regulation by transforming growth factor beta in a cutaneous T-cell lymphoma. *Blood*. Oct 15 1999;94(8):2854-61.
172. Davis TH, Morton CC, Miller-Cassman R, Balk SP, Kadin ME. Hodgkin's disease, lymphomatoid papulosis, and cutaneous T-cell lymphoma derived from a common T-cell clone. *N Engl J Med*. Apr 23 1992;326(17):1115-22. doi:10.1056/NEJM199204233261704
173. Krejsgaard T, Vetter-Kauczok CS, Woetmann A, et al. Jak3- and JNK-dependent vascular endothelial growth factor expression in cutaneous T-cell lymphoma. *Leukemia*. Oct 2006;20(10):1759-66. doi:10.1038/sj.leu.2404350
174. Woetmann A, Lovato P, Eriksen KW, et al. Nonmalignant T cells stimulate growth of T-cell lymphoma cells in the presence of bacterial toxins. *Blood*. Apr 15 2007;109(8):3325-32. doi:10.1182/blood-2006-04-017863
175. Wu CH, Yang CY, Wang L, et al. Cutaneous T-Cell Lymphoma PDX Drug Screening Platform Identifies Cooperation between Inhibitions of PI3Kalpha/delta and HDAC. *The Journal of investigative dermatology*. Feb 2021;141(2):364-373. doi:10.1016/j.jid.2020.05.110
176. Luo Y, de Gruijl FR, Vermeer MH, Tensen CP. "Next top" mouse models advancing CTCL research. *Front Cell Dev Biol*. 2024;12:1372881. doi:10.3389/fcell.2024.1372881
177. Fanok MH, Sun A, Fogli LK, et al. Role of Dysregulated Cytokine Signaling and Bacterial Triggers in the Pathogenesis of Cutaneous T-Cell Lymphoma. *The Journal of investigative dermatology*. May 2018;138(5):1116-1125. doi:10.1016/j.jid.2017.10.028
178. Jeong JH. Inducible Mouse Models for Cancer Drug Target Validation. *J Cancer Prev*. Dec 2016;21(4):243-248. doi:10.15430/JCP.2016.21.4.243

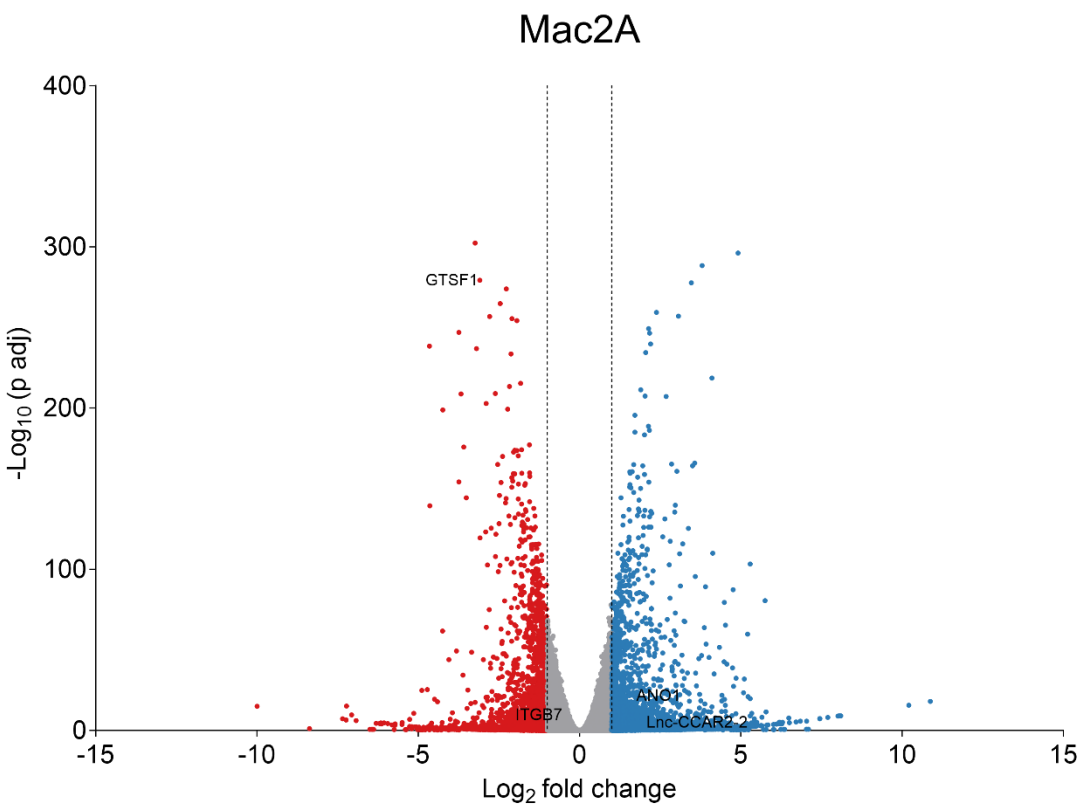
179. Ou FS, Michiels S, Shyr Y, Adjei AA, Oberg AL. Biomarker Discovery and Validation: Statistical Considerations. *J Thorac Oncol*. Apr 2021;16(4):537-545. doi:10.1016/j.jtho.2021.01.1616

Appendices



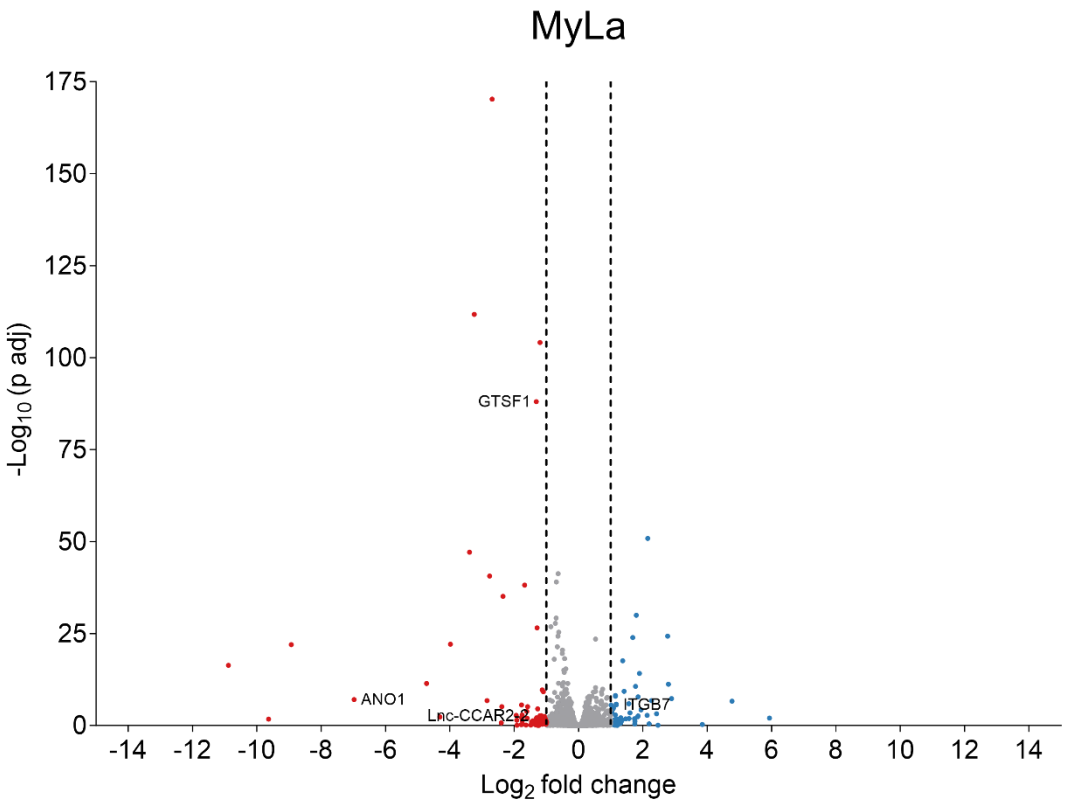
Appendix 1

Higher resolution of pathway enrichment analysis showing -Log_{10} (p-adj) of the top 50 GTSF1 interactors. Pathways from Gene Ontology (GO): Biological Processes are shown in pink and pathways from Reactome are shown in brown.



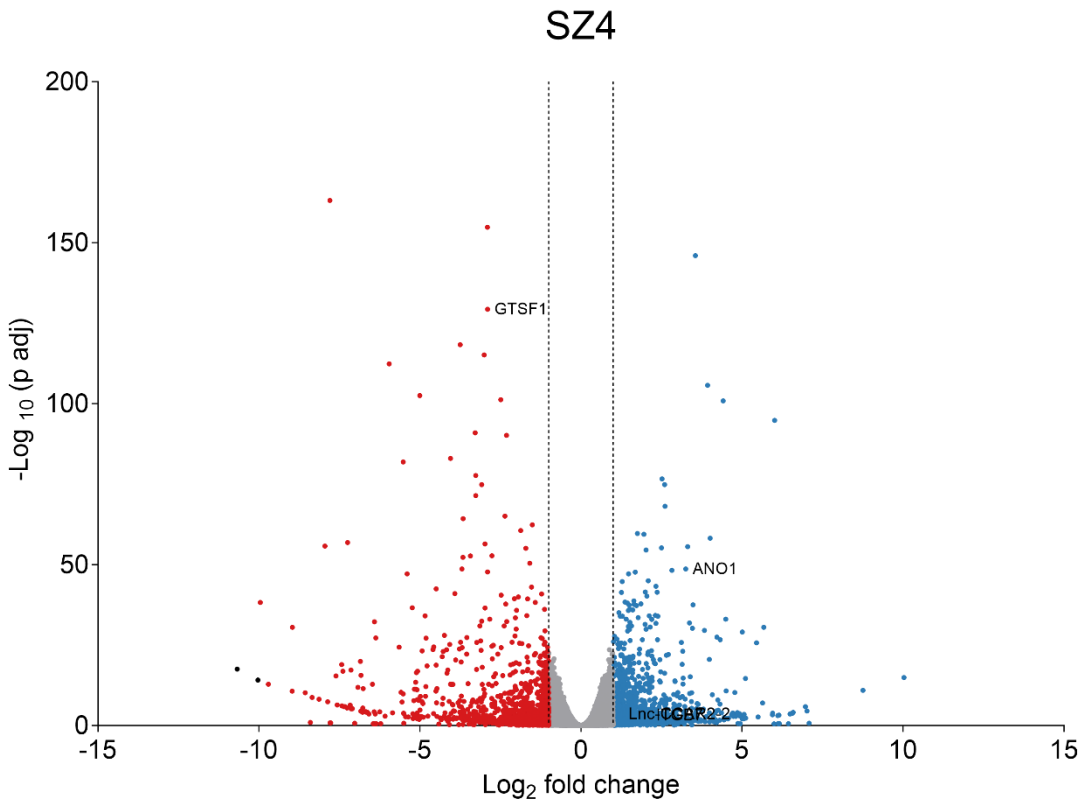
Appendix 2

Higher resolution of volcano plot of DEGs showing Log2 fold change and -Log10 (p-adj) after GTSF1 knockdown for Mac2A. Each dot represents one DEG. Upregulated genes (Log2 fold change ≥ 1) are presented in blue, downregulated genes (Log2 fold change ≤ -1) in red and genes with no significant change in grey. Highlighted are GTSF1, ANO1, ITGB7 and Lnc-CCAR2-2.



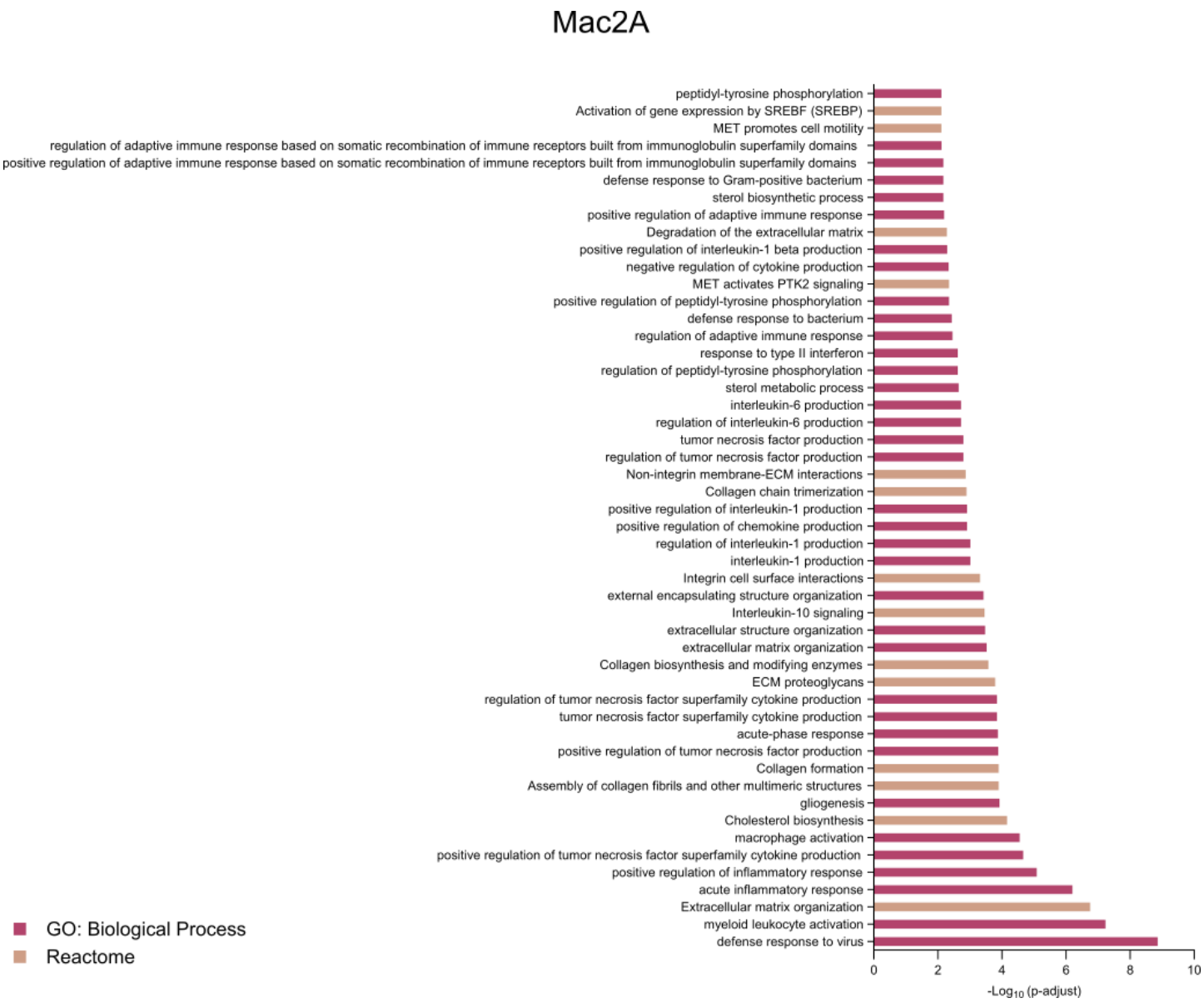
Appendix 3

Higher resolution of volcano plot of DEGs showing Log2 fold change and -Log10 (p-adj) after GTSF1 knockdown for MyLa. Each dot represents one DEG. Upregulated genes (Log2 fold change ≥ 1) are presented in blue, downregulated genes (Log2 fold change ≤ -1) in red and genes with no significant change in grey. Highlighted are GTSF1, ANO1, ITGB7 and Lnc-CCAR2-2.



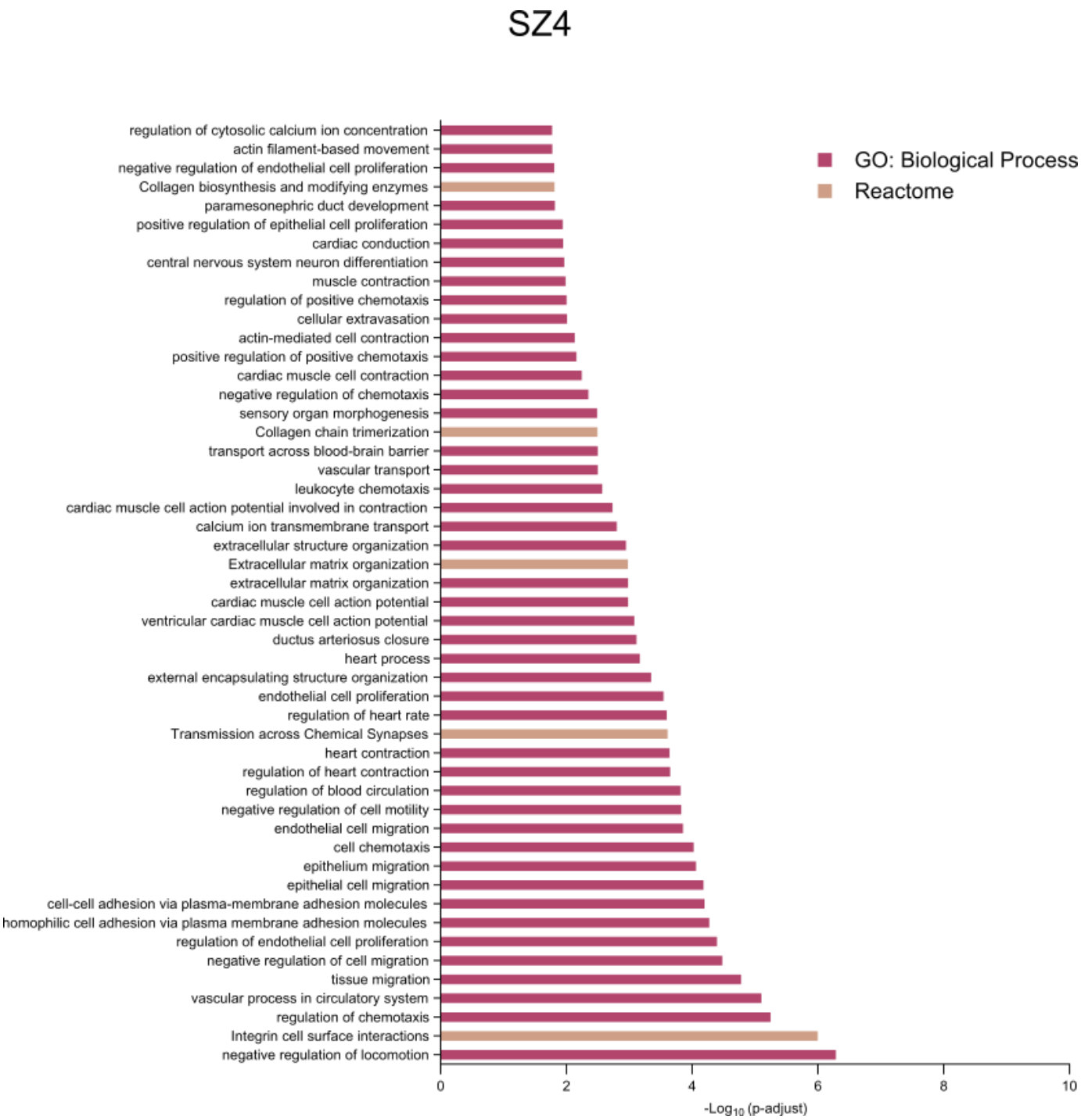
Appendix 4

Higher resolution of volcano plot of DEGs showing Log2 fold change and -Log10 (p-adj) after GTSF1 knockdown for SZ4. Each dot represents one DEG. Upregulated genes (Log2 fold change ≥ 1) are presented in blue, downregulated genes (Log2 fold change ≤ -1) in red and genes with no significant change in grey. Highlighted are GTSF1, ANO1, ITGB7 and Lnc-CCAR2-2.



Appendix 5

Higher resolution of pathway enrichment analysis showing -Log_{10} (p-adj) of DEGs after GTSF1 knockdown for Mac2A. Pathways from Gene Ontology (GO): Biological Processes are shown in pink and pathways from Reactome are shown in brown.



Appendix 6

Higher resolution of pathway enrichment analysis showing -Log_{10} (p-adj) of DEGs after GTSF1 knockdown for SZ4. Pathways from Gene Ontology (GO): Biological Processes are shown in pink and pathways from Reactome are shown in brown.