The Role of Polycomb Repressive Complexes 1 in Diffuse Large B-cell Lymphoma

Dissertation

for the award of the degree

“Doctor rerum naturalium” (Dr. rer. nat.)

of the Georg-August-Universität Göttingen

within the doctoral program Molecular Medicine

of the Georg-August University School of Science

submitted by

Débora Joppi

Born in Sorriso, Brazil

Göttingen, 2024

Thesis Advisory Committee

Prof. Dr. med. Björn Chapuy
Department of Hematology & Oncology
University Medical Center Göttingen (UMG)

Department of Hematology, Oncology and Tumor Immunology
Charité, University Medicine Berlin

Prof. Dr. Matthias Dobbelstein
Institute of Molecular Oncology
University Medical Center Göttingen (UMG)

Prof. Dr. Argyris Papantonis
Institute of Pathology
University Medical Center Göttingen (UMG)

Members of the Examination Board

Prof. Dr. med. Björn Chapuy
Department of Hematology & Oncology
University Medical Center Göttingen (UMG)

Prof. Dr. Matthias Dobbelstein
Institute of Molecular Oncology
University Medical Center Göttingen (UMG)

Further members of the Examination Board

Prof. Dr. med. Elisabeth Heßmann
Clinic for Gastroenterology and Gastrointestinal Oncology
University Medical Center Göttingen (UMG)

Prof. Dr. med. Uwe Kornak
Institute for Human Genetics
University Medical Center Göttingen (UMG)

Prof. Dr. Dieter Kube
Department of Hematology & Oncology
University Medical Center Göttingen (UMG)

“Uma mulher precisa viajar. Por sua conta, não por meio de histórias, imagens, livros ou TV. Precisa viajar por si, com seus olhos e pés, para entender o que é seu. Para um dia plantar as suas próprias árvores e dar-lhes valor. Conhecer o frio para desfrutar o calor. E o oposto. Sentir a distância e o desabrigo para estar bem sob o próprio teto. Uma mulher precisa viajar para lugares que não conhece para quebrar essa arrogância que nos faz ver o mundo como o imaginamos, e não simplesmente como é ou pode ser. Que nos faz professores e doutores do que não vimos, quando deveríamos ser alunos, e simplesmente ir ver”

"A woman needs to travel. By herself - not through stories, images, books, or TV. She needs to travel with her own eyes and feet to understand what is hers. So that one day she can plant her own trees and give them value. To know the cold to enjoy the warmth. And the opposite. To feel distance and exposure to be well under her own roof. A woman needs to travel to places she does not know to break that arrogance which makes us see the world as we imagine it, and not simply as it is or can be; which makes us professors and doctors of what we have not seen, when we should be students - simply willing to go and see"

Amyr Klink, modified

Acknowledgments

I would like to express my gratitude to my supervisor, Björn Chapuy, for giving me the opportunity to join his lab amidst a pandemic. This was a challenging journey, but I have learned so much along the way. Thank you for allowing me to grow and expand scientifically in your group and to conduct high-quality research under your guidance.

I am also grateful to my thesis advisory committee members, Professors Dobbelstein and Papantonis, for all your feedback and support throughout my PhD. A special thanks to my external examiners for agreeing to be part of my evaluation process. Thank you, Jens, for all your energy, support, and the invaluable feedback you provided to my thesis. Your motivation was crucial for me to keep going. Thank you, Sala, for your down-to-earth nature and for teaching me CUT&Tag. To all the AG Chapuy members - Luisa, Meng, Alejandro, Lisa, Vivi, Hanna, Nazli, Tobias, Rebecca, Jing, Chidimma, Shashwat – thank you for all the support, coffee-time talks, and good moments we shared. Thanks also to members of other labs from the Department of Hematology and Oncology at Charité for supporting me during this period. Sanket, Dora, Hazal, and Kaiting, those long, tiring evenings in the lab would not have been the same without you. I am also thankful for past AG Chapuy members Kirsty and Anne for all the good memories and all the others who have come and gone, leaving a bit of themselves behind. Thank you to the Javierre Lab members, Álvaro and Lucia, for all your support and feedback.

I am also deeply thankful to my family. Obrigada, mãe, por ter dado tudo de si para que eu tivesse a melhor educação e por ter me propiciado ter todas as oportunidades que tive. Acima de tudo, obrigada por ter me permitido alcançar meus sonhos. Estás comigo sempre. Obrigada aos meus irmãos Daniel, Denise e Nelsinho. Obrigada, Dani, por todo o teu apoio incondicional desde o início quando escolhi fazer farmácia, até quando me apoiaste em todos os sentidos quando fui pra Suécia. Obrigada à minha irmãzinha mais velha, Denise, por sempre se fazer presente e por ser tão prestativa - e nunca desistir de tentar fazer eu atender o celular (desculpa! agora as coisas devem melhorar). Obrigada ao meu irmão, Nelsinho, que, assim como eu, atende pouco o celular. Obrigada por me inspirar e me ensinar tanto.

Obrigada a todos os meus sobrinhos, que pouco posso ver, mas estão sempre em meus pensamentos. Durante esse doutorado, Paulo, Esther, Madalena e Joseph nasceram.
To my friends who have not given up on me during this process despite my absence and unanswered messages, thank you. My dear friends from Brazil, I miss you. Talita, Mayara, Lorena, Thayse, Kalina, Taís, Mari e Cami, vocês moram no meu coração. Minhas amigas

brasileiras em terras alemãs, Dani, Nicole, Luísa e Gabi, vocês me deram casa nesse país. Gratidão eterna. To all the friends I made since I stepped into Germany, thank you!
To my dear grandma, Esther, who left this world during my PhD writing and always asked when I would stop studying - descanse em paz, minha querida.

Finally, I want to thank my dear partner, Niklas, who became part of my life halfway through my PhD. Thank you for your unwavering support and patience, and for standing by me throughout this journey. Thank you for reminding me of the beauty in small moments when things felt overwhelming. I look forward to sharing joy with you from now on.

Table of Contents

List of Figures

List of Tables

Abbreviations

3D	Three dimensional
ABC	Activated B-cell type
AF	AlphaFold
AID	Activation-induced cytidine deaminase
AKT	Protein kinase B
AML	Acute myeloid leukemia
ANK	Ankyrin repeat region
APC	Antigen-presenting cell
BCL6	B-cell lymphoma 6
BCOR	BCL6 corepressor
BCORL1	BCL6 corepressor-like protein 1
BCR	B-cell receptor
BLIMP1	B lymphocyte-induced maturation protein-1
BSA	Bovine serum albumin
C	Cytosine
C1	Cluster 1
C2	Cluster 2
C3	Cluster 3
C4	Cluster 4
C5	Cluster 5
CBX	Chromobox
CD	Cluster of differentiation
CD40	CD40
CD40L	CD40 ligand
cDNA	Complementary DNA
CDS	Coding sequence
CHOP	Cyclophosphamide, doxorubicin, vincristine and prednisone
CHX	Cycloheximide
CLL	Chronic lymphocytic leukemia
Co-IP	Co-immunoprecipitation
COO	Cell-of-origin
CSR	Class switch recombination
Ct	Cycle threshold
DAPI	4′,6-diamidino-2-phenylindole

DepMap	The Cancer Dependency Map Project
DLBCL	Diffuse large B-cell lymphoma
DMEM	Dulbecco’s Modified Eagle’s Medium
DMSO	Dimethyl sulfoxide
DZ	Dark zone
E. coli	*Escherichia coli*
EC50	Half-maximal effective concentration
ECL	Enhanced chemiluminescence
ERK	Extracellular signal-regulated kinase
ETP-ALL	Early T-cell precursor acute lymphoblastic leukemia
EV	Empty vector
EZH2	Enhancer of zeste homolog 2
FBS	Fetal bovine serum
FDC	Follicular dendritic cell
FDR	False discovery rate
G	Guanine
GC	Germinal center
GCB	Germinal center B cell
GEO	Gene Expression Omnibus
GFP	Green fluorescent protein
GOI	Gene of interest
h	Hour(s)
H2AK119Ub	Monoubiquitination of lysine residue 119 in histone H2A
H3K27me3	Trimethylation of lysine residue 27 in histone H3
HA (tag)	Hemagglutinin
HAUSP	Herpesvirus-associated ubiquitin-specific protease
HRP	Horseradish peroxidase
ICOSL	ICOS ligand
Ig	Immunoglobulin
IGV	Integrative Genomics Viewer
IMDM	Iscove's Modified Dulbecco's Medium
IP	Immunoprecipitation
IRF4	Interferon regulatory factor 4
KDM2B	Lysine demethylase 2B
LZ	Light zone

M13r	M13 reverse
mAb	Monoclonal antibody
MAPK	Mitogen-activated protein kinase
MDS	Myelodysplastic syndrome
MEM	Minimum essential medium
MHC	Major histocompatibility complex
min	Minutes
NCoR	Nuclear receptor corepressor
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ORF	Open reading frame
PBS	Phosphate buffered saline
PcG	Polycomb group protein
PCGF	Polycomb group ring finger
PCGF1	Polycomb group ring finger 1
PCGF2	Polycomb group ring finger 2
PCGF3	Polycomb group ring finger 3
PCGF4	Polycomb group ring finger 4
PCGF5	Polycomb group ring finger 5
PCGF6	Polycomb group ring finger 6
PCR	Polymerase chain reaction
PDB	Protein data bank
PFA	Paraformaldehyde
PHC	Polyhomeotic
PI3K	Phosphoinositide 3-kinase
PNK	Polynucleotide kinase
PRC	Polycomb repressive complex
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PRDM1	PR domain zinc finger protein 1
PUFD	PCGF1 Ub-like fold discriminator
PVDF	Polyvinylidene difluoride
R	Rituximab
RING1A/B	Really interesting new gene 1A/B
RIPA	Radioimmunoprecipitation assay
RMSD	Root mean square deviation

RPMI	Roswell Park Memorial Institute Medium
RT	Room temperature
RT-qPCR	Quantitative reverse-transcription polymerase chain reaction
RYBP	Ring1- and YY1-binding protein
SCNAs	Somatic copy number alterations
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
SKP1	S-phase kinase-associated protein 1
SMRT	Silencing mediator for retinoid and thyroid hormone receptors
SOC	Super optimal broth with catabolite repression
SV	Structural variant
TAE	Tris base, acetic acid and EDTA
TBS/T	Tris-buffered saline / Tween 20®
TCR	T-cell receptor
Tfh	T follicular helper cell
UBL	Ubiquitin-like
UCH	Ubiquitin C-terminal hydrolase
USP	Ubiquitin-specific protease
USP7	Ubiquitin-specific protease 7
XBP1	X-box-binding protein 1
YAF2	YY1-associated factor 2

Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common aggressive B-cell non-Hodgkin lymphoma, known for its molecular heterogeneity. It includes transcriptomic-based subtypes, i.e. germinal center B-cell-like (GCB) and activated B-cell-like (ABC) subtypes. Recent genomic classification identified five distinct genetic subtypes of DLBCL, namely C1-C5 DLBCLs, with C3 and C4 being enriched in GCB transcriptional subtypes. C3 and C4 DLBCLs are characterized, among other factors, by alterations in chromatin regulators, including gain-of-function mutations in the Polycomb Repressive Complex 2 (PRC2) component EZH2 and in linker and core histones, respectively. PRC2 has been described as a crucial factor in normal B-cell development, and the oncogenic activation of PRC2 by gain-of-function EZH2 mutations promotes lymphomagenesis by repressing genes involved in differentiation and cell cycle control. The connection between PRC2 and Polycomb Repressive Complex 1, particularly Polycomb Repressive Complex 1.1 (PRC1.1), exists in many contexts and has also been described in naïve B cells. Given the common and well-studied *EZH2* mutations in DLBCL, I asked whether other Polycomb group proteins are also frequently mutated. In a large and informative dataset of more than 800 primary DLBCLs, I identified several subunits of the PRC1.1 complex that harbor likely loss-of-function mutations. I hypothesized that the PRC1.1 plays a tumor-suppressive role in DLBCL and that its dysfunction contributes to disease progression. Genetic analyses revealed recurrent mutations affecting PRC1.1 subunits, particularly USP7, enriched in GCB-like subtypes. Functional studies indicated that mutations in USP7 impair the half-life of USP7 and its interaction with other PRC1.1 components, such as BCOR and PCGF1, leading to destabilization of the complex. Moreover, overexpression of PCGF1 in DLBCL cell lines resulted in increased levels of H2AK119Ub, a hallmark of PRC1 activity, suggesting that PRC1.1-mediated gene repression is compromised in the presence of these mutations. Given the cooperative role of PRC1.1 with EZH2 and BCL6 in maintaining gene repression during B-cell development, along with the frequent mutations observed in PRC1.1 subunits in DLBCL, and the underexplored role of PRC1 in B-cell lymphoma, this study aimed to investigate the specific role of PRC1.1 in DLBCL.

The findings from this project support a tumor-suppressive role for PRC1.1 in DLBCL, particularly through its regulatory functions on chromatin dynamics. The disruption of PRC1.1 appears to play an important role in specific DLBCL subtypes, contributing to the deregulation of gene expression and disease progression. Ongoing studies will reveal the transcriptional targets of PRC1.1. These strategies could lead to treatments designed to prevent or reverse the oncogenic effects of PRC1.1 dysregulation, potentially improving outcomes for patients with refractory or relapsed DLBCL.

1 Introduction

1.1 The immune system

1.1.1 Innate immune system

The immune system possesses a complex, multilayered approach to combating infections and evading potential threats to the human body. The innate immune system is the initial and nonspecific line of defense. It consists of various components such as physical and chemical barriers, inflammation, the complement system, and cellular processes such as phagocytosis, mainly performed by certain immune cells, including macrophages and neutrophils[1]. The role of the innate immune system processes is to neutralize and eliminate infectious pathogens. Hence, their response is fast, providing immediate action against threats[2]. However, when a more persistent infection occurs, additional defense mechanisms, known as the adaptive immune system, arise to combat the infection[1].

1.1.2 Adaptive immune system

This targeted function is carried out by the adaptive immune system, which gradually develops following exposure to pathogens. It comprises two main types of specialized lymphocytic cells: T cells and B cells. These lymphocytes serve as effector, helper, or antibody-producing cells, collaborating to eliminate pathogens and establish long-lasting immunological memory, ensuring a quick and robust response upon re-exposure to the same pathogen[3] (***Figure 1***).

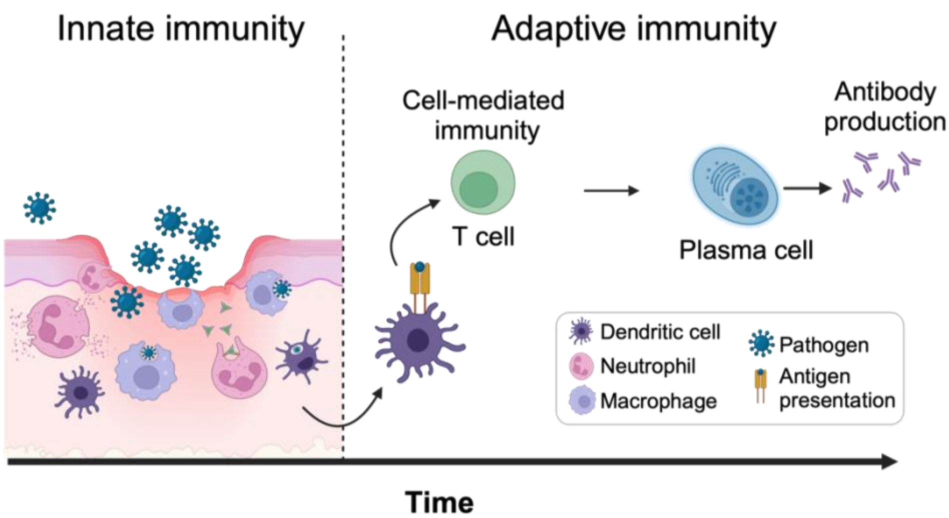

Figure 1: Evolution of innate immune responses to specialized adaptive immunity upon pathogen exposure. The illustration depicts the evolution of immune response specificity over time, starting with the early stages of immune activation at the infection site. Neutrophils and macrophages respond to the pathogen by initiating a nonspecific defense. Dendritic cells capture antigens and present them to naïve T cells (green). The activation of adaptive immunity leads to the development of highly specific T (green) and B cells (light blue), which target the pathogen more effectively. Illustration generated in BioRender by D. Joppi.

1.1.2.1 T and B lymphocytes

Both T and B lymphocytes originate from the hematopoietic stem cells in the bone marrow. T cells complete their maturation process in the thymus, where they undergo processes to promote self-tolerance, prevent autoimmunity, and develop specificity for foreign antigens. B cells, on the other hand, complete their development in the bone marrow[2]. Both types of lymphocytes express unique antigen receptors: T-cell receptors (TCRs) and B-cell receptors (BCRs), present in T and B cells, respectively. The formation of these receptors involves a process called V(D)J recombination, a random rearrangement of gene segments that occurs during the development of these cell types and creates a diversity of receptors, enabling the immune system to recognize a plethora of antigens. Both T and B cells contribute to adaptive immunity through their specialized receptors, though in distinct ways. While B cells are primarily involved in antibody production, T cells play a central role in cell-mediated immunity, which focuses on eliminating infected cells and promoting adaptive immunity.

1.1.2.2 T-cell activation and cell-mediated immunity

There are two main types of T cells: $CD8^+$-cytotoxic and $CD4^+$-helper T-cells (***Figure 2***), both of which play crucial roles in cell-mediated immunity, a defense mechanism targeting infected or abnormal cells in the body. This process requires direct contact with antigen-presenting cells (APCs), such as dendritic cells and macrophages. APCs can phagocytose pathogens and process their proteins into peptide-based antigens, which are then displayed on the surface of the APCs, bound to the major histocompatibility complex (MHC) class II molecules[1]. Additionally, intracellular antigens, such as viral proteins or abnormal proteins in tumor cells, are processed and presented by MHC class I molecules on the cell surface of nucleated cells (***Figure 2***). Upon phagocytosis, APCs migrate to lymph nodes, where the antigen presentation to T cells is conducted[1]. Cytotoxic T cells recognize MHC class I-antigen complexes, expand, leading ultimately to the targeted killing of infected cells[2]. Helper T cells, on the other hand, recognize processed antigens displayed via MHC class II on APCs, triggering a broader adaptive immune response by activating other immune cells, such as B cells[3]. This activation of B cells, induced by interaction with their cognate T cell counterpart, promotes B-cell differentiation into antibody-producing plasma cells and long-lasting memory B cells – the latter can rapidly respond upon re-exposure to the same antigen, generally without the aid of any cell[3,4] (***Figure 2***).

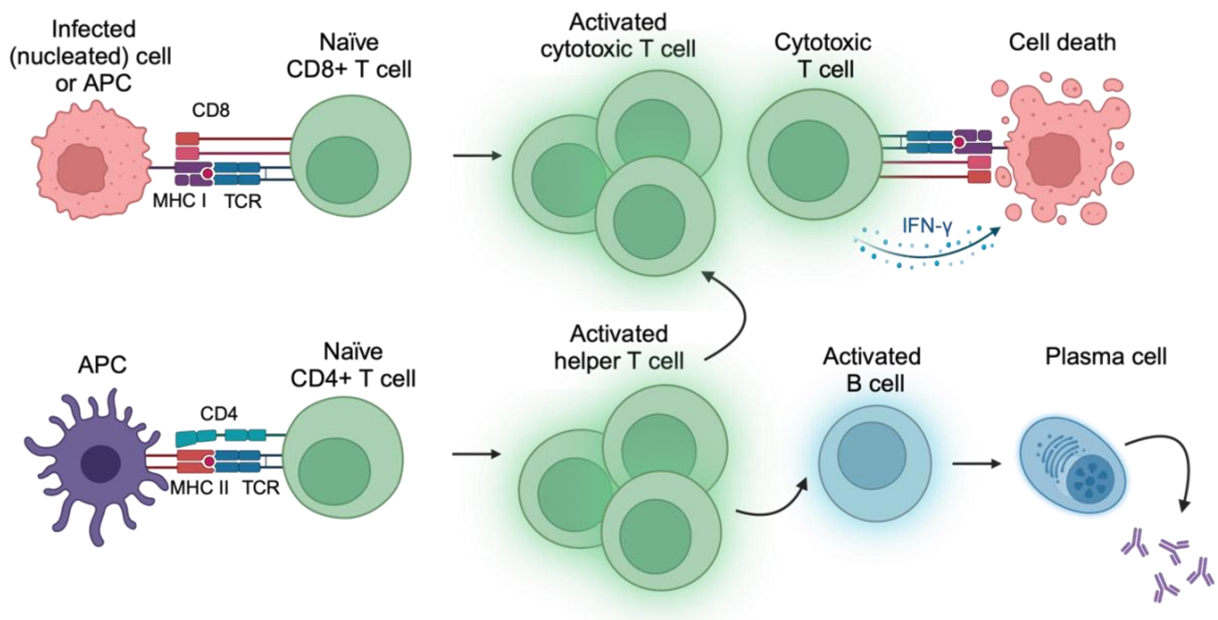

Figure 2: Activation of T cells upon presentation of antigen-MHC class I and II complexes. Naïve CD8$^+$ T cells recognize presented antigens by infected cells or antigen-presenting cells (APCs) through their major histocompatibility complex (MHC) I (in purple), leading to cytotoxic T cell activation and targeted cell killing. Naïve CD4$^+$ helper T cells recognize antigens presented by MHC class II on APCs, promoting activation of helper T cells, followed by B cell activation and differentiation into antibody-producing plasma cells. Illustration generated in BioRender by D. Joppi.

1.1.2.3 B-cell activation to antibody-mediated immunity and memory

The initial activation of B cells typically occurs when naïve B cells that have left the bone marrow encounter their specific antigen via their BCR in the peripheral blood[2,3]. After this activation, B cells are attracted by the chemokine CXCR4 and migrate to secondary lymphoid organs, such as lymph nodes, spleen, and other peripheral lymphoid tissues, i.e. tonsils and Peyer's patches. In the secondary lymphoid organs, the B cells proceed with their development. The encounter between the antigen and its specific BCR counterpart triggers a clustering of the BCR, inducing a signaling cascade within the B cell[5]. The signaling cascade involves pathways such as the phosphoinositide 3-kinase (PI3K) / protein kinase B (AKT), mitogen-activated protein kinase (MAPK) / extracellular signal-regulated (ERK), and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) pathways, that collectively promote activation, leading to changes in gene expression and consequently migration, proliferation, and differentiation of the B cell. In secondary lymphoid organs, co-stimuli provided by T follicular helper (Tfh) cells are crucial for the full activation of B cells, including processes such as class switch recombination (CSR), which allows B cells to produce different classes of antibodies that are better suited for combating specific antigens[6]. This activation is mediated through T cells binding via their Cluster of Differentiation (CD) 40 ligand (CD40L) to CD40 receptors on the surface of B cells. In addition, the CD40L-CD40 engagement is essential for promoting B cell proliferation, preventing apoptosis, and supporting the differentiation of B cells into either plasma cells or memory B cells [6–8] (***Figure* 3**). Additionally, this interaction is crucial for driving the germinal center (GC) reaction, where affinity maturation and somatic hypermutation (SHM) occur. Without this co-stimulatory signal from T cells, B cells will undergo apoptosis or become anergic, preventing full activation and differentiation. This tightly regulated mechanism prevents B cells from producing antibodies against pathogens that are not considered harmful to the organism.

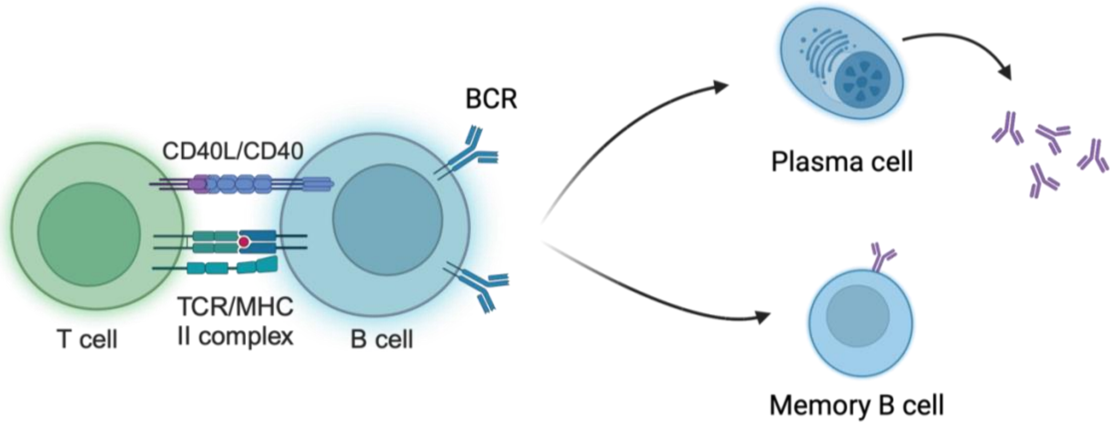

Figure 3: Activation of B cells upon interaction with T cells leading to B cell differentiation. The interaction between T and B cells is crucial for B cell activation and occurs through the binding of CD40 ligand (CD40L) on the T cell to the CD40 receptor on the B cell, and the engagement of the T cell receptor (TCR) with the major histocompatibility complex (MHC II), presented by the B cell. This interaction promotes the differentiation of B cells into antibody-secreting plasma cells or memory B cells, as shown on the right side. Plasma cells secrete antibodies that contribute to the immune response. Illustration generated in BioRender by D. Joppi.

Following activation, B cells can have distinct differentiation fates based on the initial affinity of their BCR for the antigen and the intensity of CD40-CD40L binding. If the BCR has high affinity for the antigen (1), the B cells may differentiate into short-lived plasma cells that quickly produce large amounts of immunoglobulin (Ig) M, initiating a rapid immune response aimed at neutralizing the pathogen[3]. (2) If the BCR affinity is lower, some B cells may bypass the GC altogether, differentiating into GC-independent memory B cells, which can provide immune memory without undergoing the full GC process [9,10]. Otherwise, B cells can go through another fate to elicit a long-term and robust immune response: (3) the generation of plasma and memory B cells in structures called GCs, where the selection of competent clones and CSR and affinity maturation of antibodies occurs[2,3,10]. The production of antibodies by B cells promotes the elimination of the pathogens, which can take place through neutralization, where antibodies bind to extracellular antigens and prevent them from interacting with host cells; complement-dependent cytotoxicity, where antibodies activate the complement system, leading to the infected cell destruction; antibody-dependent cellular phagocytosis, in which macrophages or neutrophils phagocytose pathogens bound by antibodies; antibody-dependent cellular cytotoxicity, where natural killer cells recognize infected cells flagged by antibodies and release of cytotoxic molecules to induce their death.

1.1.3 The formation of the germinal center

To produce high-affinity antibodies and long-lasting memory B cells, B lymphocytes within GCs experience affinity maturation of the variable regions within the BCR. The GCs are transient and dynamic structures that are formed in secondary lymphoid organs upon B cell activation. In the GC, B cells proliferate rapidly and undergo selection[4,11]. Histologically, this compartment is divided into the dark zone (DZ), where B cells clonally expand and SHM occurs. The dark zone is separated from the light zone (LZ)[11], where antigen-exposed B cells are selected based on antigen affinity with the help of Tfh and follicular dendritic cells (FDCs) (***Figure 4***)[11,12]. In the DZ compartment, rapidly dividing B cells, called centroblasts, experience SHM, a process that introduces mutations into the variable regions of Ig genes through the deamination of cytosine to uracil mediated by activation-induced cytidine deaminase (AID)[13]. AID activity leads to DNA damage, with subsequent repair processes that result in somatic changes in the DNA that yield changes to the BCR affinity. Centroblasts express characteristically high levels of B-cell lymphoma 6 (BCL6)[14], a master regulator of the GC program, as well as CXCR4, which directs their migration to the DZ[15]. BCL6 represses DNA damage responses and prevents premature differentiation, while CXCR4 mediates the retention of centroblasts in the DZ, where they undergo further proliferation and mutation. In the LZ, centrocytes express high levels of CXCR5, which is required for their localization in this compartment, and decreased levels of CXCR4[15]. They upregulate CD86 and ICOS ligand (ICOSL), molecules critical for interactions with Tfh cells[16]. Centrocytes with successful affinity maturation of the BCR are more proficient at recognizing the presented antigen by FDCs and are hence clonally selected (positive selection)[17]. Centrocytes that effectively present MHC Class II-antigen complexes to Tfh cells receive essential survival and differentiation signals, including IL-4 and CD40L signaling, which are also critical for CSR[16,18]. During CSR, centrocytes switch the antibody isotype they produce, allowing the immune response to be tailored more effectively against the specific type of pathogen. Those centrocyte clones without proper binding do not receive proper co-stimulation and consequently undergo apoptosis (negative selection). Antigen-exposed B-cells with low-affinity BCRs that were not negatively selected either experience further rounds of affinity maturation or differentiate into memory B cells. In fact, a subset of MYC-expressing LZ B cells re-enters the DZ for additional cycles of mutation and selection, thereby further refining antibody specificity[16,19,20]. This cyclical process of proliferation, mutation, and selection ensures a robust and highly specific immune response. The selected centrocytes then further differentiate either into antibody-secreting plasma cells or memory B cells (***Figure 4***)[21].

The fate of B cells during these maturation processes is governed by the expression of key transcription factors: B lymphocyte-induced maturation protein 1 (BLIMP1, as known as PRDM1) promotes plasma cell differentiation and drives the expression of genes essential for antibody secretion[22], which is also done by X-box-binding protein 1 (XBP1). BCL6, on the other hand, is crucial for the formation of the GC and helps maintain germinal center B cells by inhibiting PRDM1 and enhancing the survival of these cells within the GC[14]. Additionally, interferon regulatory factor 4 (IRF4) plays a dual role, inducing plasma cell differentiation by downregulation of BCL6. High levels of IRF4, in fact, lead to a shift towards the plasma cell fate, whereas lower levels are associated with memory B cell development[23]. Another important factor for the initiation of the GC is MYC. Together with BCL6, MYC plays an essential role in B-cell development: without them, there is no formation of GCs. Moreover, the balance between the expression of MYC and BCL6 is pivotal in determining whether a B cell becomes a memory cell or differentiates into a plasma cell[16,20,24].

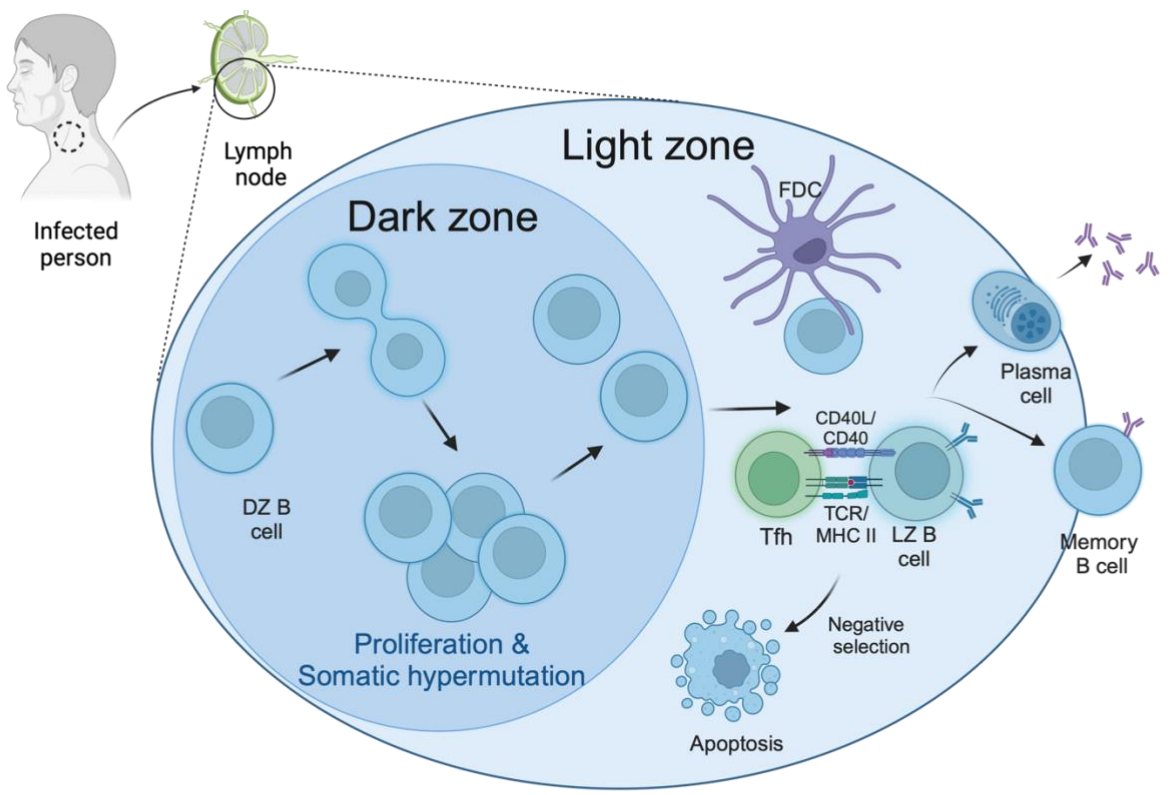

Figure 4: Dynamics of B-cell maturation and selection in the germinal center. The cartoon illustrates the anatomy of the germinal center, in this case located within a lymph node of an infected person, highlighting the distinct roles of the dark and zones. In the dark zone (DZ), B cells undergo rapid proliferation and somatic hypermutation to diversify their B-cell receptors. DZ B cells migrate to the light zone (LZ), where they interact with follicular dendritic cells (FDCs) and T follicular helper (Tfh) cells via CD40L/CD40 signaling. LZ B cells with high affinity receive survival signals, differentiating into plasma or memory B cells, whereas low-affinity or autoreactive B cells are subjected to apoptosis. Illustration generated in BioRender by D. Joppi.

1.1.4 The germinal center is the origin of B-cell lymphoma

The very molecular processes that drive the dynamic maturation of B cells also expose them to risks of genomic instability. The high proliferative rate, SHM, and AID-mediated DNA damage[24] place a significant burden on the genomic integrity of GC[24]. This, coupled with the suppression of DNA damage checkpoints by BCL6 and other regulatory factors, renders B cells more susceptible to malignant transformation. When the balance between proliferation, differentiation, and apoptosis is lost, dysregulated GCs can emerge, creating a favorable environment for oncogenesis. The failure of mechanisms that typically prevent uncontrolled growth or repair faulty DNA can lead to the accumulation of genetic mutations that drive the development of lymphomas[24,25].

One such B-cell lymphoma is diffuse large B-cell lymphoma (DLBCL), which exhibits a highly heterogeneous landscape[25,26]. The molecular heterogeneity of DLBCL frequently stems from genetic alterations that impact key regulatory genes essential for the normal function of GC B cells[24,27] and this is further reflected in the diverse clinical manifestations of the disease.

1.2 Diffuse large B-cell lymphoma

1.2.1 Clinical characteristics

The most common type of Non-Hodgkin lymphoma is DLBCL, accounting for 30 to 40% of all lymphoma. Patients usually present with large lymph nodes and systemic symptoms such as fever, night sweats, and weight loss[28]. As approximately 40% of DLBCL present with disease outside secondary lymphoid organs (also known as “extranodal”), this lymphoma might present with other symptoms that are associated with the physiological function of the perturbed organ. The incidence of DLBCL increases with age, more frequently occurring in elderly patients with a median age of 65 at the time of diagnosis. If left untreated, DLBCL is rapidly fatal, highlighting the importance of prompt diagnosis and treatment.

1.2.2 Treatment options

The standard treatment for DLBCL includes a combination of rituximab (R), an anti-CD20 monoclonal antibody (mAb), and CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone) chemotherapy. Recently, vincristine has been replaced by polatuzumab-vedotin, an antibody drug conjugate of the anti-CD79B antibody with a directly linked synthetic vinca alkaloid, monomethyl auristatin E (MMAE)[29]. This regimen, known as R-CHOP or, most recently, Pola-R-CHP, has significantly improved survival rates, with approximately 65% of patients achieving a cure[28,30]. However, due to the clinical and molecular heterogeneity of DLBCL, a substantial number of patients either relapse after initial treatment or have refractory disease that does not respond to standard therapy. This variability highlights the critical need for better biomarkers and more effective, targeted therapies to improve outcomes for patients, particularly those with high-risk diseases or who are resistant to first-line treatment[31].

1.2.3 Transcriptional heterogeneity in DLBCL

The clinical complexity of DLBCL arises from its molecular heterogeneity. DLBCL can be classified based on transcriptional differences in several ways[25,26,32,33], with one of the most widely used being the cell-of-origin (COO) classification [32]. In this classification, DLBCL is classified based on the transcriptional similarity of the tumor to normal GC B-cells or *in vitro* stimulated peripheral B-cells into two distinct subtypes: (1) germinal center B-cell (GCB)-type and (2) activated B-cell (ABC)-type DLBCL[32]. The GCB-type DLBCL arises from light zone germinal B-cells re-entering the dark zone (***Figure 5***) and has a more favorable prognosis following R-CHOP frontline treatment. GCB DLBCL is characterized by the expression of genes typically active in germinal B cells, such as *BCL6*[14], and by frequent genetic alterations in Enhancer of zest homolog 2 (*EZH2)*, a methyltransferase that plays a key role in epigenetic regulation in normal B cells and lymphomagenesis[34]. In contrast, ABC-type DLBCL originates from a late stage of the GC reaction, where B cells are committed to plasmablastic or memory B-cell differentiation (***Figure 5***)[32]. This subtype is associated with a worse prognosis and resistance to traditional therapeutic regimens[35]. The ABC-type DLBCL is characterized by chronic active BCR signaling and NF-κB pathway, frequently resulting from genetic alterations in *MYD88*, *CARD11*, and *CD79B*[36,37].

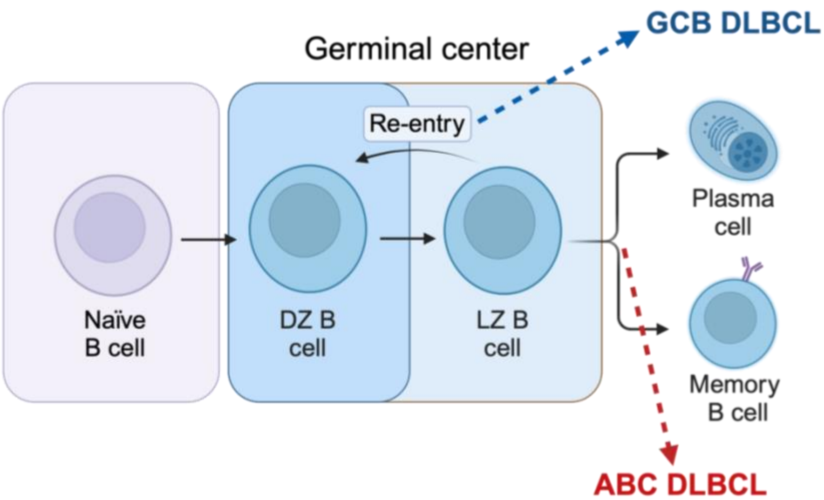

Figure 5: Cell-of-origin classification of DLBCL. Cartoon illustrating B cell differentiation within the germinal center, showing GCB-type DLBCL arising from light zone B cells prone to dark zone re-entry, and ABC-type DLBCL originating from post-germinal center B cells differentiating into memory or plasma cells. Illustration generated in BioRender by D. Joppi.

1.2.4 Genetic heterogeneity in DLBCL

The classification of DLBCL into GCB and ABC subtypes has provided meaningful insights into the clinical characteristics and treatment response of this disease. However, it is now evident that the genetic profile of DLBCL has more degrees of complexity that are worth looking at. In the past years, advances in genomic profiling have led to the development of more detailed molecular classifications that capture the genetic diversity of DLBCL[25,38]. One of these classifications, DLB*class,* has identified five distinct genetic clusters (C1-C5) based on integrated analyses of mutations, somatic copy number alterations (SCNAs), and structural variants (SVs)[25] (***Figure 6A***). Each cluster is associated with specific genetic alterations and dysregulated signaling pathways, offering deeper insights into the mechanisms of lymphomagenesis and allowing for more precise prognostic and treatment predictions. Additionally, an independent study identified within the LymhGen classification initially three genetic subtypes i.e. MCD, BN2, and EZB, which overlap with three of the clusters (C1~BN2, C3~EZH2, C5~MCD)[26]. In a reanalysis, the same group later identified two additional clusters, ST2 and A53, which share similarities with the other two clusters (C2~A53, C4~ST2)[38]. The favorable risk C1 DLBCLs and the BN2-type DLBCLs share molecular similarities, including their enrichment in transcriptional ABC-type DLBCLs, frequent *NOTCH2* mutations, and *BCL6* translocations. The COO independent C2 DLBCLs share some similarities with A53 subtype and are defined by *TP53* biallelic inactivation (typically one allele lost and one mutated), genomic stability, and a distinct but unfavorable outcome.

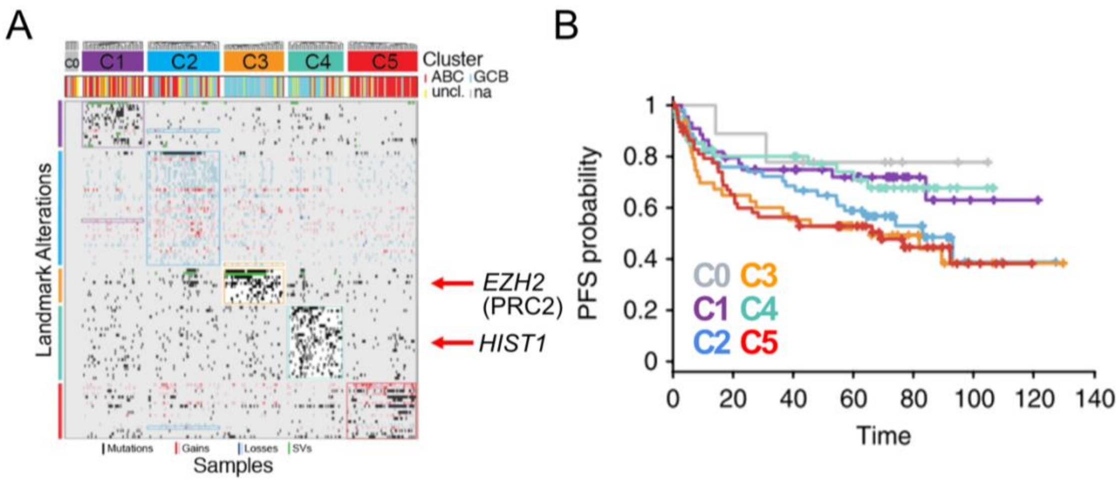

Figure 6: Molecular subtypes of DLBCL based on frequent mutations, somatic copy number alterations, and structural variants. (A) Heatmap showing frequent mutations, somatic copy number alterations (gains and losses), and structural variants across DLBCL samples, grouped into clusters C1 to C5. Alterations in epigenetic components such as *EZH2* and *HIST1* are highlighted. (B) Kaplan-Meier survival analysis indicating progression-free survival (PFS) probability for each genetically defined DLBCL subtype. Data are obtained and modified from Chapuy *et al*., 2018.

The unfavorable C5 DLBCLs share similarities with the MCD type, which is highly enriched in ABC-type DLBCL, and typically harbors mutations in *MYD88* and *CD79B* genes, leading to chronic BCR signaling and NF-κB activation. Finally, C3 and C4 DLBCLs are GCB-type-enriched clusters, overlapping with EZB and ST2 subtypes, respectively. C3 exhibits frequent BCL2 translocation and frequent mutations in chromatin modifiers, such as *EZH2*, *CREBBP, KMT2D,* and *EP300*, while C4 includes patients with mutations in histones such as histone H1, as well as alterations in NF-κB or RAS–JAK pathways[25]. In terms of clinical implications, C4 has a favorable outcome, while C3 has a poor prognosis (***Figure 6B***).

1.2.4.1 Alterations in epigenetic components in C3 and C4 DLBCLs

The GCB-type predominant C3 DLBCLs are characterized by cells that are epigenetically locked into the GC transcriptional program, maintaining uncontrolled proliferation upon DNA damage and suppression of the expression of differentiation-promoting genes. These cells are thus not able to exit the GC reaction and differentiate. This cluster frequently exhibits mutations in chromatin modifiers, such as *EZH2*[25], the catalytic component of the Polycomb repressive complex (PRC) 2 (PRC2), which is essential for the GC reaction in B cells[39]. PRCs, in general, are multiprotein complexes that repress gene expression through histone modifications. PRC2, through EZH2, deposits methyl groups on lysine 27 of histone H3 (H3K27me3), resulting in chromatin compaction and repressed gene expression[40]. Gain-of-function mutations in *EZH2* are common, particularly in DLBCL[41], which enhances the enzymatic activity of this protein and leads to increased chromatin compaction in C3-DLBCL[42,43]. Additional frequently mutated genes include *KMT2D*, *CREBBP*, and *EP300*, which are involved in histone modification, leading to diverse effects on chromatin structure and therefore gene expression[34]. The C4 cluster, on the other hand, is distinguished by its genetic heterogeneity and includes a broad spectrum of mutations, particularly in histones. The mutations are thought to affect the structure of the chromatin and to potentiate an increase in mutation rates, by the exposure of the DNA to continuous AID activity[25]. These diverse chromatin modifier mutations suggest that the tumors from these clusters exhibit varying levels of epigenetic dysregulation, contributing to their heterogeneous nature and potential resistance to standard therapies[44,45].

1.3 Epigenetics and epigenomics

1.3.1 Epigenetic modifications

The malleability of the genome to express genes and to give rise to different cells upon division, such as a highly proliferative GC B cell or a quiescent memory B cell, relies on the epigenetic modification of the genome. This is particularly important in a dynamic and genomic damage-inducing environment such as the GC reaction, in which B cells are locked in a certain epigenetic state to allow genetic alterations to occur during the physiologic B-cell maturation processes, SHM and CSR as detailed above[46]. These epigenetic modifications include DNA methylation and histone modifications, which ultimately lead to changes in the expression of genes they are close to or related to[47].

1.3.2 DNA methylation

DNA methylation is one of the best-characterized epigenetic modifications. It typically acts as a repressive mark through the addition of a methyl group to the 5' position of cytosine (C) residues, primarily within CpG islands, regions of the DNA that have a high concentration of C nucleotides followed by guanine (G), often occurring in promoter regions of genes. This modification leads to gene silencing by preventing the binding of transcription factors or by recruiting proteins that compact chromatin, thereby reducing gene expression[48–50].

1.3.3 Histone modifications

Histone modifications are post-translational modifications of histone proteins that regulate chromatin structure and function. These include methylation, acetylation, ubiquitination, and phosphorylation, and occur on the amino (N)-terminal tails of histone proteins, playing a pivotal role in determining whether chromatin is in an active or repressive state. Modifications such as methylation serve as signals for various chromatin-associated proteins, often referred to as "readers". Other enzymes, known as "writers" and "erasers", add or remove these modifications, respectively, to dynamically regulate gene expression[51,52]. Among these modifications, some marks are particularly associated with gene repression, and one of the primary epigenetic modifier complexes responsible for establishing and maintaining these repressive states are the PRCs[53].

1.3.4 Polycomb repressive complexes

PRCs are essential epigenetic regulators for preserving cellular identity, governing developmental processes, and preventing undesirable gene expression. PRCs were first described in the fruit fly *Drosophila melanogaster*, where they were found to repress the expression of genes responsible for regulating the development of anatomical structures in this organism[54,55]. More specifically, they suppress homeotic genes named *Hox* genes in certain regions of the fly during the developmental stage, ensuring their expression is limited to specific body segments[56]. In mammals, PRCs have been described to play important roles in regulating stem cell maintenance and differentiation processes, as well as in the development and progression of various types of cancer[53]. The two main types of PRCs are PRC1 and PRC2, each with distinct compositions and functions.

1.3.4.1 Polycomb Repressive Complex 2 (PRC2)

The subunits of PRC2 are essential for its gene-silencing function through histone modifications. The components include the catalytic subunits, the N-methyltransferases EZH2 or its less common paralog EZH1, both of which mediate the trimethylation of lysine 277 of the tail of histone H3 (H3K27me3). This repressive histone mark plays a pivotal role in maintaining chromatin in a condensed, transcriptionally inactive configuration[40]. EZH2 is the primary catalytic subunit in most actively proliferating tissues, while EZH1 has been shown to partially compensate for the loss of EZH2, particularly in quiescent cells[57]. The H3K27me3 epigenetic modification is associated with transcriptional silencing and chromatin compaction, playing a critical role in maintaining the repressed state of genes involved in development and cellular differentiation. In B cells, and DLBCL in particular, H3K27me3 is deposited to silence genes involved in pathways that control cell cycle regulation, apoptosis, and differentiation, such as *CDKN1A* and PR domain zinc finger protein 1 *(PRDM1)*[39,45]. The repression of these genes helps B cells maintain their proliferative capacity within the GC while preventing premature differentiation. However, in the context of DLBCL, aberrant deposition of H3K27me3 due to mutations in the Y641 residue, particularly the gain-of-function Y641F/N mutations, in the catalytic C-terminal SET domain of EZH2, leads to the silencing of tumor suppressor genes and other regulatory genes that are crucial for surveillance of uncontrolled cell growth[41]. These mutations not only enhance the methyltransferase activity of EZH2 but also significantly alter the chromatin architecture of DLBCL cells, leading to a more compact and repressive chromatin state that promotes lymphomagenesis[42,58,59]. The relevance of PRC2 in DLBCL and lymphomagenesis has been extensively investigated, promoting the development of specific EZH2 inhibitors for clinical use[60,61]. One such EZH2 inhibitor, tazemetostat, was approved by the FDA for patients with relapsed or refractory follicular lymphoma and is currently undergoing clinical trials[62]. Alterations in PRC2 and an increase in H3K27me3 have also been reported in patients with prostate, sarcoma, colon, and breast carcinomas, and they point to an oncogenic function of PRC2[63]. However, a tumor suppressor role has also been described for EZH2 and PRC2, with inactivating *EZH2* mutations occurring in myelodysplastic syndrome (MDS)[64] and early T-cell precursor acute lymphoblastic leukemia (ETP-ALL)[65].

1.3.4.2 Polycomb Repressive Complex 1 (PRC1)

In addition to the repressive H3K27me3 mark, another critical repressive modification is the monoubiquitination of histone H2A at lysine 119 (H2AK119Ub) catalyzed by PRC1[66]. Compared to PRC2, PRC1 exhibits greater structural and functional diversity[67]. The PRC1 family consists of canonical and non-canonical compositions, which in total form six multiprotein complexes that catalyze H2AK119Ub through the E3 ubiquitin ligases Really interesting new gene 1 (RING1) B (RING1B) and its paralog RING1A[67,68]. The nomenclature of the PRC1 subcomplex depends on the Polycomb group RING finger (PCGF) subunit it contains, ranging from PCGF1 to PCGF6 (PRC1.1 to PRC1.6)[69]. PRC1.2 and PRC1.4, as well as PRC1.3 and PRC1.5, are closely related and can overlap, sharing core components. The canonical and non-canonical PRC1s assemble into diverse subcomplexes (

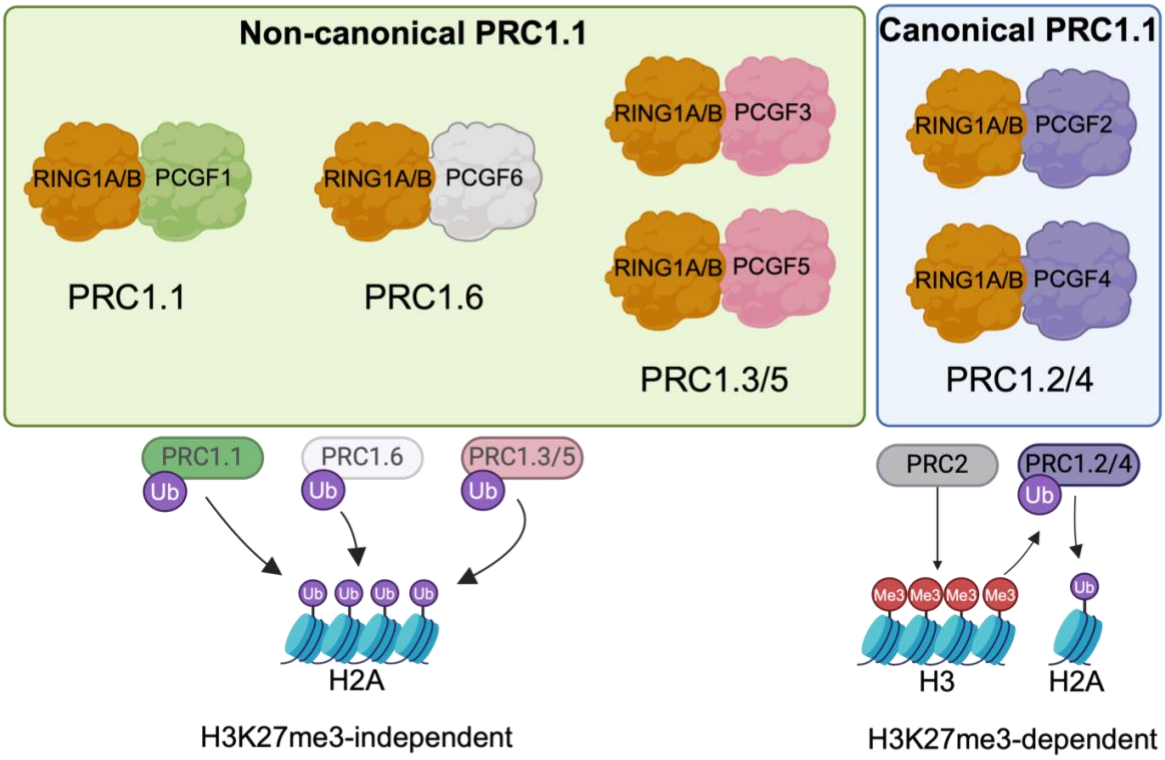

Figure 7).

Figure 7: PCGF family proteins form canonical and non-canonical PRC1 complexes with PRC2-dependent and independent H2AK119Ub-mediated deposition. The non-canonical PRC1 complexes contain RING1A/B plus PCGF1 (PRC1.1), PCGF6 (PRC1.6), PCGF3 or PCGF5 (PCGF3 or PCGF5), which independently mediate ubiquitination of histone H2AK119 (Ub). The canonical PRC1 complex includes, apart from RING1A/B, PCGF2 or PCGF4 (PRC1.2/4), which are dependent on PRC2-mediated H3K27me3 marks for recruitment and subsequent H2AK119Ub deposition. PRC2 catalyzes the methylation of histone H3 at lysine 27 (H3K27me3). Illustration generated in BioRender by D. Joppi.

PRC1 and PRC2 can be functionally interconnected, with PRC2-mediated H3K27me3 serving as a recruitment signal for canonical PRC1 to the chromatin[70]. Upon recruitment, canonical PRC1 catalyzes the addition of ubiquitin to H2AK119, further reinforcing the repressive chromatin structure by inhibiting transcriptional elongation and maintaining gene silencing. This cooperative action between PRC2 and PRC1 contributes to the stable and heritable repression of target genes, playing a crucial role in both normal B cell function and the pathogenesis of DLBCL when dysregulated[39,59,66]. However, PRC1 can also deposit H2AK119Ub independently of PRC2 activity, a function facilitated by non-canonical PRC1 subcomplexes[71].

1.3.4.3 Canonical PRC1

Canonical PRC1 is composed of chromobox proteins (CBX), RING1A/B, the PCGF proteins PCGF2 (also known as MEL18) and PCGF4 (also known as BMI1), and polyhomeotic (PHC) proteins[69]. CBX proteins are critical for binding to the H3K27me3 mark, facilitating the recruitment of PRC1 to condensed chromatin[72]. PCGF2 and PCGF4 help to scaffold the complex and modulate the activity of PRC1 by contributing to the stability and specificity of the complex[69]. PHC proteins play a role in the higher-order chromatin compaction that PRC1 mediates[73].

1.3.4.4 Non-canonical PRC1

Non-canonical PRC1 subcomplexes, on the other hand, do not require PRC2-mediated H3K27me3 for their recruitment to chromatin. Instead, they are recruited by DNA-binding factors, allowing PRC1 to not only function independently of PRC2 but also recruit it, thus establishing a feedback loop that reinforces gene silencing. The basic structure of non-canonical PRC1 contains RING1A/B, PCGF Proteins (PCGF1, PCGF3, PCGF5, PCGF6) and the mutually exclusive RYBP/YAF2[70].

1.3.4.5 Non-canonical PRC1.1

PRC1.1 is a specific variant of non-canonical PRC1 characterized by its unique subunit composition and PRC2-independent chromatin recruitment activity for the ubiquitination of histone H2A. Loss-of-function mutations in non-canonical PRC1 components can lead to the derepression of critical oncogenes, resulting in uncontrolled cell proliferation and tumor development[74]. Specifically, dysregulation of PRC1.1 subunits, including BCOR, KDM2B, and RING1A/B, has been associated with the development of several cancers, such as acute myeloid leukemia (AML) and MDS[74–76]. PRC1.1 is composed of the following protein subunits:

Ubiquitin-specific peptidase or protease (USP7), also known as herpesvirus-associated ubiquitin-specific protease (HAUSP), is a deubiquitinase from the ubiquitin-specific protease (USP) family first described by Everett and colleagues in 1997[77]. USP7 has been observed to have many different target proteins, which are dependent or independent of its deubiquitinase activity[78]. In PRC1.1, USP7 interacts with other core components, such as KDM2B, PCGF1, BCOR, and RING1B[79], being described as essential for keeping the integrity and stability of the complex by deubiquitinating PRC1.1 subunits such as RING1B, therefore preventing its polyubiquitylation and consequent degradation by the proteasome machinery[80]. Pharmacological inhibition of USP7 results in the disassembly of PRC1.1 and loss of the complex's ability to bind to chromatin in leukemia, where it is considered an oncogene[81]. USP7 displays a duality in its role across different cancer types, functioning either as a tumor suppressor or an oncogene depending on the context. This ambivalence is largely dictated by its interactions with regulators of cellular homeostasis. For instance, USP7 is known to stabilize both the tumor suppressor p53 and its negative regulator MDM2. While the stabilization of p53 by USP7 usually happens under stress, by deubiquitinating and stabilizing MDM2, USP7 indirectly promotes p53 degradation, contributing to tumor progression[82]. In contrast, in certain cancers such as AML, USP7 stabilizes factors such as DNMT1 and PTEN, further driving leukemogenesis[83]. Given its context-dependent functions, USP7 is considered a potential target for therapeutic interventions aimed at restoring the balance of protein stability, with many inhibitors having been developed, such as FT671 and P5091[84,85]. However, in B-cell lymphoma, the role of USP7 remains poorly understood, warranting further research to elucidate its contribution to DLBCL and its potential as a therapeutic target.

PCGF1 is an exclusive component of PRC1.1 and is central for the assembly of this complex[67]. It interacts with RING1A/B, the catalytic subunit responsible for the ubiquitination of histone H2A, as well as with BCOR and its paralog BCORL1, and then KDM2B, upon BCOR or BCORL1 binding to PCGF1[86,87]. PCGF1 requires interaction with RING1A/B to bind the complex to target sites[88]. Disruptions or dysregulation of PCGF1 can compromise the function of PRC1.1, resulting in the loss of gene silencing[89].

BCOR (BCL6 co-repressor) is a fundamental subunit of PRC1.1 that has an important function in chromatin regulation. BCOR is a transcriptional corepressor that has also been shown to interact with other corepressors, such as BCL6. In a mutually exclusive way, BCOR competes for binding with silencing mediator for retinoid and thyroid hormone receptors (SMRT) and nuclear receptor corepressor (NCoR), which form other complexes with BCL6 that also repress target genes[90,91]. Mutations in *BCOR* have been frequently observed in hematological cancers, including AML, MDS, and chronic lymphocytic leukemia (CLL)[75,92,93]. *BCOR* mutations are often loss-of-function, leading to the impairment of BCOR's repressive function, resulting in the dysregulation of genes that promote uncontrolled proliferation and cancer progression[94,95]. In AML, *BCOR* mutations are found in 4-6% of cases, being associated with poor prognosis by disrupting normal blood cell development and promoting leukemia[92,96,97]. In MDS, mutations are linked to unfavorable outcomes, especially in higher-risk subtypes, by driving genomic instability and disrupting cell differentiation, increasing the risk of transformation to AML[75,98]. In CLL, on the other hand, *BCOR* mutations are less common but contribute to aggressive disease and treatment resistance, often co-occurring with mutations like *NOTCH1* or *TP53*[99]. Thus, *BCOR* usually has mutations disrupting its function which are important drivers of malignancy in hematological cancers[76,99].

RING1A/B are ubiquitin ligases that act as the catalytic subunits of PRC1.1. These proteins are responsible for monoubiquitinating histone H2A at lysine 119. RING1B is the most frequently studied of the two and is essential for maintaining chromatin in a silenced state[89]. Disruption of RING1A/B impairs the repressive function of PRC1 and has been associated with defects in the maintenance and differentiation of stem cells[68,69].

The Lysine Demethylase 2B (KDM2B) is a histone-modifying enzyme that selectively removes methyl groups from histone H3 at lysine 36 (H3K36me2), a mark typically associated with active gene transcription. KDM2B has been described as a crucial subunit for PRC1.1, as it recognizes and binds to unmethylated CpG islands at target gene promoters and recruits the PRC1.1 complex to these sites, enabling the complex to mediate chromatin compaction and gene repression[100]. Disruptions in KDM2B have been implicated in leukemias and other hematological neoplasias[101]. Loss of KDM2B activity results in the accumulation of H3K36me2, which impairs the proper recruitment of PRC1.1 to chromatin. This leads to the aberrant activation of gene expression programs. Such dysregulation promotes uncontrolled cell proliferation and survival, thereby contributing to oncogenesis. Moreover, mutations or other genetic alterations in *KDM2B* disrupt the balance between gene activation and repression, favoring pathways that drive tumorigenesis[69].

The S-phase kinase-associated protein 1 (SKP1) associates with KDM2B and forms a stable subcomplex with this protein. SKP1 is part of the minimal four-component structure required for the assembly of PRC1.1, which suggests an important role of SKP1 for the stability of the complex.

Ring1- and YY1-binding protein (RYBP) and YY1-associated factor 2 (YAF2) are paralogs with high homology that can have functionally redundant roles. They are found exclusively in non-canonical PRC1, and both proteins can stimulate the E3 ubiquitin ligase activity of PRC1 complexes, by binding specifically to H2AK119Ub1, which then recruits more RYBP/YAF2-PRC1 complexes to target sites in the chromatin.

1.4 Aims of the study

Given the centrality of the oncogenic-activated PRC2 complex in the biology of DLBCL and the interconnection of the PRC1 and PRC2 complexes, this study aims to investigate the specific role of the PRC1 complexes in the development and progression of DLBCL. A special focus will be on the PRC1.1 complex as a critical epigenetic regulator involved in gene repression and its dysregulation in DLBCL by proposing the following specific aims:

1. **Investigate the frequency of genetic alterations in Polycomb group proteins in DLBCL** by identifying the frequency and distribution of genetic alterations in primary DLBCL patient cohorts.

2. **Examining the expression levels of PRC1.1 subunits** in DLBCL patients and the normal B-cell ontogeny to assess any dysregulation.

   To evaluate the RNA of PRC1.1 subunits in DLBCL and B cells, I will analyze RNA-seq data from DLBCL patient cohorts and healthy individuals, respectively. Publicly available immunohistochemistry images will be used to localize some of these subunits in lymph node sections from healthy controls. Additionally, I will assess PRC1.1 protein levels by Western blot.

3. **Characterizing the molecular impact of PRC1.1 mutations** by analyzing their functional consequences in DLBCL, focusing on:

   a. *The epigenetic regulatory functions of PRC1.1*

   b. *The impact of mutations on the interaction between PRC1.1 subunits.*

   c. *The effect of these mutations on protein stability.*

4. **Functional modeling of the molecular implications of PRC1.1 dysfunction** in DLBCL by determining whether inhibiting or restoring PRC1.1 subunits can modulate oncogenic pathways.

5. ***In silico* modeling of PRC1.1 complex and impact of its mutations**

By addressing these objectives, this study seeks to shed light on the understanding of the physiological contribution of PRC1.1 during B-cell ontogeny and its role following genetic deregulation in DLBCL. A key focus is to understand why PRC1.1 disruption takes place in DLBCL and determine whether restoring or compensating for its loss of PRC1.1 offers therapeutic benefits. These strategies could lead to treatments designed to prevent or reverse the oncogenic effects of PRC1.1 dysregulation, potentially improving outcomes for patients with refractory or relapsed DLBCL.

2 Materials & Methods

2.1 Material

2.1.1 Chemicals and reagents

Table 1: Chemicals and reagents used

Reagent	Company
4′,6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich, Burlington, Massachusetts, USA
Agar	Carl Roth, Karlsruhe, Germany
Agarose	Carl Roth, Karlsruhe, Germany
alamarBlue™	BioRad, Hercules, California, USA
Ampicillin, sodium salt	Carl Roth, Karlsruhe, Germany.
Bolt™ LDS Sample buffer (4x)	Invitrogen, Waltham, Massachusetts, USA
Bolt™ SDS MES running buffer (20x)	Invitrogen, Waltham, Massachusetts, USA
Bolt™ sample reducing agent (10x)	Invitrogen, Waltham, Massachusetts, USA
Bolt™ transfer buffer (20x)	Invitrogen, Waltham, Massachusetts, USA
Bovine serum albumin (BSA)	Carl Roth, Karlsruhe, Germany
Chloroform, ≥99.5%	Honeywell, Charlotte, North Carolina, USA
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, Burlington, Massachusetts, USA
Dulbecco's Modified Eagle's Medium (DMEM)	Gibco, Waltham, Massachusetts, USA
Enhanced Chemiluminescence (ECL) - Immobilon® Crescendo buffer ultra western horseradish peroxidase (HRP) substrate	Merck Millipore, Burlington, Massachusetts, USA
Ethanol, ≥99%	Carl Roth, Karlsruhe, Germany
Fetal bovine serum (FBS)	Gibco, Waltham, Massachusetts, USA
Gel loading dye purple (6x)	New England Biolabs, Ipswich, Massachusetts, USA
GeneRuler 1kb DNA Ladder	Thermo Scientific, Waltham, Massachusetts, USA
Halt™ protease and phosphatase inhibitor cocktail (100x)	Thermo Scientific, Waltham, Massachusetts, USA
HEPES buffer 1M	Gibco, Waltham, Massachusetts, USA
Hexadimethrine bromide (Polybrene)	Sigma-Aldrich, Burlington, Massachusetts, USA
Iscove's Modified Dulbecco's Medium (IMDM)	Gibco, Waltham, Massachusetts, USA
Isopropanol, 99.5%	Thermo Fisher Scientific, Waltham, Massachusetts, USA
Kanamycin	Carl Roth, Karlsruhe, Germany
L-glutamine	Sigma-Aldrich, Burlington, Massachusetts, USA
Methanol, 99%	Carl Roth, Karlsruhe, Germany
Midori green advance DNA stain	Nippon Genetics Europe, Düren, Germany
NEBNext® high-fidelity 2X PCR master mix	New England Biolabs, Ipswich, Massachusetts, USA
Nuclease and DNA-free free water	Genaxxon, Ulm, Germany
Opti-minimum essential medium (MEM)	Gibco, Waltham, Massachusetts, USA

PageRuler™ Plus prestained protein ladder	Thermo Scientific, Waltham, Massachusetts, USA
Paraformaldehyde (PFA) in phosphate buffered saline (PBS), 4%	Süsse, Gudensberg, Germany
DPBS	Gibco, Waltham, Massachusetts, USA
Penicillin/streptomycin (10000U/mL penicillin, 10mg/mL streptomycin)	Gibco, Waltham, Massachusetts, USA
Pierce Anti-HA magnetic beads	Thermo Scientific, Waltham, Massachusetts, USA
Pierce Protein G magnetic beads	Thermo Scientific, Waltham, Massachusetts, USA
Poly-D-Lysine (0.1mg/mL)	Gibco, Waltham, Massachusetts, USA
Polyethylenimine (PEI)	Sigma-Aldrich, Burlington, Massachusetts, USA
Propidium Iodide	Sigma-Aldrich, Burlington, Massachusetts, USA
Restore™ PLUS Western Blot Stripping Buffer	Thermo Scientific, Waltham, Massachusetts, USA
Radioimmunoprecipitation Assay (RIPA) lysis buffer (10x)	Merck Millipore, Burlington, Massachusetts, USA
Roswell Park Memorial Institute Medium (RPMI 1640)	Gibco, Waltham, Massachusetts, USA
Sodium chloride	Carl Roth, Karlsruhe, Germany
Super optimal broth with catabolite repression (SOC) medium	Invitrogen, Waltham, Massachusetts, USA
T4 DNA ligase reaction buffer (10x)	New England Biolabs, Ipswich, Massachusetts, USA
T4 polynucleotide kinase buffer (10x)	New England Biolabs, Ipswich, Massachusetts, USA
TransIT-293 Transfection Reagent	Mirus Bio, Madison, Wisconsin, USA
Tris-base	Carl Roth, Karlsruhe, Germany
Tris-HCl	Carl Roth, Karlsruhe, Germany
TRIzol™ reagent	Invitrogen, Waltham, Massachusetts, USA
TrypanBlue solution	Sigma-Aldrich, Burlington, Massachusetts, USA
Trypsin-EDTA, 0.05%	Gibco, Waltham, Massachusetts, USA
Tryptone	Carl Roth, Karlsruhe, Germany
Tween® 20	Serva Electrophoresis, Heidelberg, Germany
Yeast extract	Carl Roth, Karlsruhe, Germany

Table 2: Buffers and solutions

Buffer	Composition	Purpose
1x RIPA lysis buffer	1x RIPA lysis buffer 1x proteinase and phosphatase inhibitor in ddH_2O	Total protein isolation
Agar plates	2% (w/v) agar in lysogeny broth medium	Bacterial growth
Agarose gel	0.8% (w/v) agarose 1x Midori green advance DNA stain in ddH_2O	DNA separation
Antibody dilution solution	0.5% (w/v) BSA 0.01% (w/v) NaN_3	Immunoblotting

	in TBS/T	
Cao lysis buffer	20mM HEPES, pH 7.9 250mM NaCl 20mM b-glycerophosphate 5mM 4-nitrophenyl phosphate-Na_2 salt 1mM EDTA 0.1% NP-40 1% Glycerol in ddH_2O	Subcellular fractionation
Freezing medium	10% (v/v) DMSO in fetal bovine serum (FBS)	Freezing cells
Lysogeny broth (LB) medium	10 g/L tryptone 5 g/L yeast extract 5 g/L NaCl, pH 7.0 in ddH_2O	Bacterial growth
MES SDS Running Buffer	1x Bolt™ MES SDS Running Buffer (20x) in ddH_2O	Immunoblotting
Tris base, acetic acid and EDTA (TAE) buffer	40 mM Tris, pH 7.5-9.0 20 mM glacial acetic acid 1 mM EDTA in ddH_2O	DNA separation in agarose gel electrophoresis
Transfer buffer	50 mL Bolt™ Transfer Buffer 100 mL Methanol (99%) in 850mL ddH_2O	Protein transfer
Tris-buffered saline / Tween 20® (TBS/T)	137 mM NaCl 20 mM Tris-HCl (pH 7.6) 0.1% (v/v) Tween® 20 in ddH_2O	Polyvinylidene difluoride (PVDF) membrane washing; dilution of antibodies for immunoblotting

Table 3: Media composition

Medium	Composition
DMEM + 10% FBS (D10)	10% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in DMEM
DMEM + 30% FBS (D30)	30% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in DMEM
IMDM + 10% FBS (I10)	10% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in IMDM

IMDM + 20% FBS (I20)	20% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in IMDM
RPMI 1640 + 10% FBS (R10)	10% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in RPMI 1640
RPMI 1640 + 20% FBS (R20)	20% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 10 mM HEPES in RPMI 1640
RPMI 1640 + 10% FBS + pyruvate (R10p)	10% FBS 2 mM L-Glutamine 100 U/mL Penicillin 100 µg/mL Streptomycin 1 mM pyruvate 10 mM HEPES in RPMI 1640

2.1.2 Consumables

Table 4: Consumables used

Consumable	Company
Cell culture flask, T-25 and T-75 for suspension cells	Sarstedt, Nümbrecht, Germany
Cell culture plate for suspension cells (6, 12, 24, 48 and 96 wells)	Sarstedt, Nümbrecht, Germany
Cell culture plates for adherent cells (6 wells)	Sarstedt, Nümbrecht, Germany
Cell strainer pluriStrainer Mini (70 µm)	pluriSelect Life Science, Leipzig, Germany
Centrifuge tubes	Sarstedt, Nümbrecht, Germany
CryoPure tubes (2 mL)	Sarstedt, Nümbrecht, Germany
Culture tubes (14 mL)	Greiner Bio-One, Kremsmünster, Austria
Filtertips (10, 20, 100, 200 and 1000 µL)	Sarstedt, Nümbrecht, Germany
Flow cytometry tubes (5 mL)	Sarstedt, Nümbrecht, Germany
Gel-load pipette tip (200 µL)	Greiner Bio-One, Kremsmünster, Austria
Glass coverslips (24x55 mm)	Epredia, Kalamazoo, Michigan, USA
Hybond® P PVDF membrane (0.45 µm)	Cytiva, Marlborough, Massachusetts, USA
Parafilm® M Sealing Film	Bemis, Neenah, Wisconsin, USA
PCR tubes (200 µL)	Sarstedt, Nümbrecht, Germany
Safeseal reaction tube (0.5, 1.5, 2 and 5 mL)	Sarstedt, Nümbrecht, Germany
Screw cap tubes (15 and 50 mL)	Sarstedt, Nümbrecht, Germany
Serological pipette (2, 5, 10 and 25 mL)	Sarstedt, Nümbrecht, Germany
Superfrost® microscope slides (1 mm)	Epredia, Kalamazoo, Michigan, USA

Syringe filters (0.45 µm)	Carl Roth, Karlsruhe, Germany
Syringes with Luer Lock (50 mL)	B. Braun, Melsungen, Germany
Whatman® gel blotting paper GB002	Cytiva, Marlborough, Massachusetts, USA

2.1.3 Laboratory devices

Table 5: Devices used

Device	Product name	Company
Agarose gel electrophoresis system	Mupid™-One Advance	Nippon Genetics Europe, Düren, Germany
Benchtop centrifuge	5810 R	Eppendorf, Hamburg, Germany
Chemiluminescence imaging system	ECL Chemocam	Intas Science Imaging Instruments, Göttingen, Germany
CO_2 Incubator	CB260	Binder, Tuttlingen, Germany
Cytocentrifuge	Shandon Cytospin III	Thermo Scientific, Waltham, Massachusetts, USA
Flow cytometer	BD FACSAria	Becton Dickinson, Franklin Lakes, New Jersey, USA
Freezer, -20 °C	GNP 2756 Nofrost	Liebherr, Bulle, Switzerland
Freezer, -80 °C	Ultra-Low temperature freezer	Panasonic, Kadoma, Osaka, Japan
Fridge, 4 °C	Öko Super Glass Line	Liebherr, Bulle, Switzerland
Gel imager	Quantum CX5	Vilber Lourmat, Collégien, France
Microcentrifuge	5425 R	Eppendorf, Hamburg, Germany
Microvolume Spectrophotometer	NanoDrop™ 2000	Thermo Fisher Scientific, Waltham, MA, USA
Multichannel pipette (8-channel)	Research plus 10-100µL	Eppendorf, Hamburg, Germany
PCR thermal cycler	C1000 Touch Thermal Cycler	BioRad, Hercules, California, USA
pH meter	pH 211	Hanna Instruments, Woonsocket, Rhode Island, USA
Pipette aid	Pipette boy 2	Integra Biosciences, Zizers, Switzerland
Pipettes (2.5, 10, 100, 1000µL)	Research plus series	Eppendorf, Hamburg, Germany
Platform shaker	Unimax 2010	Heidolph Instruments, Schwabach, Germany
Protein gel electrophoresis chamber system	Minigel tank	Thermo Scientific, Waltham, Massachusetts, USA
qPCR cycler	StepOnePlus™	Thermo Scientific, Waltham, Massachusetts, USA
Spectrofluorometer	Synergy LX	BioTek Instruments, Winooski, Vermont, USA
Spectrophotometer	800 TS	BioTek Instruments, Winooski, Vermont, USA
Thermomixer	Thermomixer C	Eppendorf, Hamburg, Germany
Vortex mixer	Vortex-Genie® 2	Scientific Industries, Bohemia, New York, USA
Water bath	GFL 1083	GFL, Vaughan, Ontario, Canada

2.1.4 Cell lines

Table 6: Human cell lines used in the study

Cell line	Cell type	Medium
B-JAB	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
Balm-3	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
CTB1	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
DB	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
Farage	Primary mediastinal B-cell lymphoma	RPMI 1640 + 10% FBS
HBL1	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS + 1mM pyruvate
HEK293T/17	Embryonic kidney	DMEM + 10% FBS
HT	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
Karpas 422	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
OCI-Ly1	Diffuse large B-cell lymphoma	IMDM + 10% FBS
OCI-Ly10	Diffuse large B-cell lymphoma	IMDM + 20% FBS
OCI-Ly18	Diffuse large B-cell lymphoma	IMDM + 10% FBS
OCI-Ly3	Diffuse large B-cell lymphoma	IMDM + 20% FBS
OCI-Ly4	Diffuse large B-cell lymphoma	IMDM + 20% FBS
OCI-Ly7	Diffuse large B-cell lymphoma	IMDM + 10% FBS
Pfeiffer	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
RI-1 (RIVA)	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
SU-DHL-10	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
SU-DHL-4	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
SU-DHL-5	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
SU-DHL-6	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
TMD8	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
Toledo	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
U2932	Diffuse large B-cell lymphoma	RPMI 1640 + 10% FBS
U2940	Primary mediastinal B-cell lymphoma	RPMI 1640 + 10% FBS

All cell lines used in this study were provided by our collaboration partner M. Shipp (Dana-Farber Cancer Institute, Boston, Massachusetts, USA) as short tandem repeats (STR)-profiled cells. From here on, OCI-Ly1, OCI-LY10, OCI-LY3, OCI-LY4 and OCI-LY7 will be referred to as “LY1”, “LY10”, “LY3”, “LY4” and “LY7”, respectively. Similarly, Karpas 422 will be referred to as “K422”, B-JAB as “BJAB” and SU-DHL-10. SU-DHL-4, SU-DHL-5 and SU-DHL-6 will be abbreviated as “DHL10”, “DHL4”, “DHL5” and “DHL6”, respectively, and HEK293T/17 as “HEK293T”.

2.1.5 Cloning and expression vectors

Table 7: Cloning vectors used

Backbone plasmid	Bacterial resistance	Mammalian selection	Origin	Plasmid number
Gateway™ pDONR™221	Kanamycin	NA	Invitrogen, Waltham, Massachusetts, USA	12536017

lentiGuide-Puro	Ampicillin	Puromycin	Addgene, Cambridge, Massachusetts, USA	52963
pLX_307	Ampicillin	Puromycin	Broad Institute, Cambridge, Massachusetts, USA	pLX_307
pLX_312	Ampicillin	Green fluorescent protein (GFP)	Broad Institute, Cambridge, Massachusetts, USA	pLX_312

From here on, pLX_307 and pLX_312 containing vectors are referred to as pLX307 and pLX312, respectively. NA, nonapplicable.

Table 8: Plasmids used for lentiviral packing system

Plasmid	Bacterial resistance	Function	Origin	Plasmid number
pMD2.G	Ampicillin	Envelope plasmid	Addgene, Cambridge, Massachusetts, USA	12259
psPAX2	Ampicillin	Packaging plasmid	Addgene, Cambridge, Massachusetts, USA	12260

Table 9: Open reading frames plasmid used

Open reading frame (ORF) plasmid	Bacterial resistance	Origin	Plasmid number
pCR8/GW/TOPO PCGF1	Spectinomycin	Addgene, Cambridge, Massachusetts, USA	69770
pQFlag-USP7 WT puroR	Ampicillin	Addgene, Cambridge, Massachusetts, USA	46751

Table 10: Expression plasmid produced and used in this study

All the plasmids below were cloned by Débora Joppi, unless otherwise indicated.

Expression plasmid	Lentiviral backbone	Gene insert	Tag
pLX307_GFP (*)	pLX307	GFP	V5
pLX307_PCGF1	pLX307	Wild-type PCGF1 ORF	V5
pLX307_USP7_WT	pLX307	Wild-type USP7 ORF	HA
pLX312_CTRL (**)	pLX312	Insertion of a non-coding sequence between the chloramphenicol resistance and the *ccdB* gene	None
pLX312_USP7_WT	pLX312	Wild-type USP7 ORF	HA
pLX312_Y243D	pLX312	USP7 ORF with Y243D point mutation	HA
pLX312_I660K	pLX312	USP7 ORF with I660K point mutation	HA
pLX312_T730S	pLX312	USP7 ORF with T730S point mutation	HA
pLX312_Y1056H	pLX312	USP7 ORF with Y1056H point mutation	HA
pLX307_ΔTRAF_ΔUCH	pLX312	USP7 with deletions of ΔTRAF and ΔUCH domains	HA

pLX307_ΔUBL1-5	pLX312	USP7 with deletions of ΔUBL1-5 domains	HA

(*) The pLX307_GFP expression plasmid was cloned by Björn Chapuy in the Shipp Lab at Dana-Farber Cancer Institute in Boston, Massachusetts, USA.
(**) The pLX312_CTRL plasmid was originally obtained as pDONR221_CTRL from Konstandina Isaakidis from the Keller Group of Hematology, Oncology, and Cancer Immunology at the Benjamin Franklin Campus of Charité, Berlin. The LR recombination reaction and further plasmid isolation were performed by Débora Joppi.

2.1.6 Kits

Table 11: Kits used

Kit	Company
DC Protein Assay	Bio-Rad, Hercules, California, USA
Gateway® BP clonase II	Invitrogen, Waltham, Massachusetts, USA
Gateway® LR clonase II	Invitrogen, Waltham, Massachusetts, USA
GoTaq® qPCR kit	Promega Corporation, Madison, Wisconsin, USA
mPAGE®TurboMix Bis-Tris Gel Casting	Sigma-Aldrich, Burlington, Massachusetts, USA
NucleoBond® Xtra Midi EF	Macherey-Nagel, Düren, Deutschland
Phusion High-Fidelity DNA Polymerase	Thermo Scientific, Waltham, Massachusetts, USA
QiaPrep Spin Miniprep Kit	Qiagen, Hilden, Germany.
RevertAid® first strand cDNA synthesis kit	Thermo Scientific, Waltham, Massachusetts, USA
Wizard® SV Gel and PCR Clean-Up	Promega Corporation, Madison, Wisconsin, USA

2.1.7 Antibodies

Table 12: Primary antibodies used for immunoblotting

Antibody	Species, isotype	Concentration	Catalog Number	Company
BCL6	Rabbit, IgG	1:1000	sc-858	Santa Cruz Biotechnology, Dallas, Texas, USA
BCOR	Rabbit, IgG	1:1000	63972	Cell Signaling Technology, Danvers, Massachusetts, USA
GAPDH	Mouse, IgG1 κ	1:1000	sc-47724	Santa Cruz Biotechnology, Dallas, Texas, USA
H2AK119Ub	Rabbit, IgG	1:1000	8240	Cell Signaling Technology, Danvers, Massachusetts, USA
HA-probe	Rabbit, IgG	1:1000	sc-805	Santa Cruz Biotechnology, Dallas, Texas, USA
HAUSP (USP7)	Mouse, IgG1 κ	1:700	sc-137008	Santa Cruz Biotechnology, Dallas, Texas, USA

Lamin B1	Rabbit, IgG	1:1000	12586	Cell Signaling Technology, Danvers, Massachusetts, USA
PCGF1	Mouse, IgG1 κ	1:700	sc-515371	Santa Cruz Biotechnology, Dallas, Texas, USA
RING1B	Rabbit, IgG	1:1000	13069	Cell Signaling Technology, Danvers, Massachusetts, USA
RYBP	Mouse, IgG2b	1:700	sc-374235	Santa Cruz Biotechnology, Dallas, Texas, USA
V5-Tag	Rabbit, IgG	1:1000	13202	Cell Signaling Technology, Danvers, Massachusetts, USA

Table 13: Antibodies used for immunoprecipitation

Antibody	Species, isotype	Total amount	Catalog Number	Company
HAUSP (USP7)	Mouse, IgG1 κ	3µg	sc-137008	Santa Cruz Biotechnology, Dallas, Texas, USA
V5	Mouse, IgG2a	3µg	sc-271944	Santa Cruz Biotechnology, Dallas, Texas, USA
Normal IgG1	Mouse, IgG1 κ	3µg	sc-3877	Santa Cruz Biotechnology, Dallas, Texas, USA
Normal IgG2a	Mouse, IgG2a	3µg	sc-3878	Santa Cruz Biotechnology, Dallas, Texas, USA

Table 14: Secondary antibodies used for immunoblotting

Antibody	Species, isotype	Concentration	Catalog Number	Company
HRP-conjugated anti-mouse	Goat, IgG	1:10000	115035003	Jackson ImmunoResearch Laboratories, West Grove, Pennsylvania, USA
HRP-conjugated anti-rabbit	Goat, IgG	1:10000	111035144	Jackson ImmunoResearch Laboratories, West Grove, Pennsylvania, USA

2.1.8 Primer sequences

All primers listed in the tables below are shown in the 5' to 3' orientation.

Table 15: Cloning primers

Oligonucleotide	Sequence	Purpose
USP7_del1_RV_attB2	GGGGACAACTTTGTACAAGAAAGCTGGG CTCAAGCGTAGTCTGGGACGTCGTATGG GTAATCTTGTGGAAATCCCATG	Amplification of ΔUBL1-5 *USP7*
USP7_FW_attB1	GGGGACAACTTTGTACAAAAAAGCAGGCT GCCATGAACCACCAGCAG	Amplification of WT *USP7*
USP7_RV_attB2	GGGGACAACTTTGTACAAGAAAGCTGGG CTCAAGCGTAGTCTGGGACGTCGTATGG GTAGTTATGGATTTTAATGGCCTTTTCAAG	Amplification of WT *USP7*
USP7_UCHdel_FW_attB1	GGGGACAACTTTGTACAAAAAAGCAGGCT GCCATGATCAGGGAATCAAAACTGAG	Amplification of ΔTRAF ΔUCH *USP7*

Table 16: Sequencing primers

Oligonucleotide	Sequence	Purpose
EF1alpha	TCAAGCCTCAGACAGTGGTTC	Sequencing
M13	TGTAAAACGACGGCCAG	Sequencing
M13 reverse (M13r)	CAGGAAACAGCTATGAC	Sequencing
USP7_E572_FW	AGAGGACCAGTTTTGTGGC	Sequencing
USP7_V918_RV	ACACACCCATGCTTGTCT	Sequencing

Table 17: Real-time PCR primers

Oligonucleotide	Sequence	Purpose
HA_tag_qPCR FW	TACCCATACGACGTCCCAGAC	Real-time PCR
HA_tag_qPCR RV	AGCGTAGTCTGGGACGTCGTATG	Real-time PCR
USP7 qPCR FW	GCGGCAAAATGGTGTCCTAT	Real-time PCR
USP7 qPCR RV	TTGTCCCCATCGAGCTGTTC	Real-time PCR
GAPDH FW	GATTCCACCCATGGCAAATTC	Real-time PCR
GAPDH RV	TGATTTTGGAGGGATCTCGCTC	Real-time PCR

Table 18: Guide RNA sequences used for CRISPR/Cas9-mediated gene editing of *USP7*

Oligonucleotide	Sequence	Purpose
AAVS1 guide RNA FW (*)	CACCGGGGGCCACTAGGGACAG GAT	Non-targeting control (AAVS1 site)
AAVS1 guide RNA RV (*)	AAACATCCTGTCCCTAGTGGCCC CC	Non-targeting control (AAVS1 site)
USP7 guide RNA 1 FW	CACCGAGATGTATGATCCCAAAA CG	CRISPR targeting Exon 8 of *USP7*

USP7 guide RNA 1 RV	AAACCGTTTTGGGATCATACATCTC	CRISPR targeting Exon 8 of *USP7*
USP7 guide RNA 2 FW	CACCGGGCAACCTTTCAGTTCACTG	CRISPR targeting Exon 17 of *USP7*
USP7 guide RNA 2 RV	AAACCAGTGAACTGAAAGGTTGCCC	CRISPR targeting Exon 17 of *USP7*
USP7 guide RNA 3 FW	CACCGGGCAGTAGAACAGCTCGATG	CRISPR targeting Exon 15 of *USP7*
USP7 guide RNA 3 RV	AAACCATCGAGCTGTTCTACTGCCC	CRISPR targeting Exon 15 of *USP7*

(*) Designed by and obtained from Joji Shimono from the Chapuy Lab from the Department of Hematology and Oncology in Charité, Berlin.

Table 19: Primers used for site-directed mutagenesis

Oligonucleotide	Sequence	Purpose
USP7_I660K FW	ATAATGAAAACCCTTGGACAAaATTCCTGGAAACAGTTGATCC	Site-directed mutagenesis
USP7_I660K RV	CACTGAGCTCAATCATTGTTTTATTGCCGTCG	Site-directed mutagenesis
USP7_T730S FW	TGACAGAGCAGGATTTATTCAAGATAgTAGCCTTATCCTCT	Site-directed mutagenesis
USP7_T730S RV	CACATAACTGGGAGCAAGTCACGTATTTTACAG	Site-directed mutagenesis
USP7_Y1056H FW	CACCAGTACATAAATGAAGACGAGcATGAAGTAAATTTGAAAGACTTTG	Site-directed mutagenesis
USP7_Y1056H RV	TCGGCCCATCATTACAATTGCAAATTTAAACTTCTC	Site-directed mutagenesis
USP7_Y243D FW	GGGGGATGATTCGTCTAAAAGCGTCCCT	Site-directed mutagenesis
USP7_Y243D RV	TCGGTTGGCATCATGTcCACAGCCTTTCGTAGC	Site-directed mutagenesis

Lowercase letters indicate the altered nucleotides introduced for mutagenesis.

2.1.9 Bacteria strains

Table 20: Bacteria strains used for the amplification of entry and expression plasmids

Bacterial strain	Purpose	Company
E. coli One Shot™ TOP10	Amplification of viral expression plasmids	Invitrogen, Waltham, Massachusetts, USA
E. coli One Shot™ Stbl3™	Amplification of entry plasmids	Invitrogen, Waltham, Massachusetts, USA

2.1.10 Drugs used

Table 21: Drugs used in human cell lines from the study

Drug	Aim	Concentrations used	Company
FT671	USP7 inhibition	48.8 nM – 50 µM	MedChemExpress, Monmouth Junction, New Jersey, USA

Cycloheximide (CHX)	Translation inhibition	10 µg/mL	Clini Sciences, Nanterre, France
Puromycin	Selection of puromycin-resistant cells	2 µg/mL	InvivoGen San Diego, California, USA

2.1.11 Publicly available data

2.1.11.1 Publicly available genomics data sets

Table 22: Publicly available data used in this study

Publication	**Data type**	**Access number (GEO or dbGaP database)**
Brune *et al*., 2008	RNA sequencing	Refer to Brune *et al*., 2008
Chapuy *et al*., 2018	RNA sequencing	GSE98588
	Whole exome sequencing	phs000450
Schmitz *et al*., 2018	RNA sequencing	Refer to Schmitz *et al*., 2018
	Whole exome sequencing	phs001444.v2.p1

2.1.11.2 Publicly available protein structures

PCGF1	**ID (Protein Data Bank [PDB] or AlphaFold database [AF])**		**Domains covered**
PCGF1	AF	AF-Q9BSM1-F1-v4	Full length
PCGF1/BCOR	PDB	4HPL	RAWUL (PCGF1)
			PUFD (BCOR)
PCGF1/BCOR/KDM2B	PDB	8HCU	RAWUL (PCGF1)
			PUFD (BCOR)
			Leucine-rich repeats (LRRs) (KDM2B)
USP7	AF	AF-Q93009-F1-v4	Full length
	PDB	2F1W	TRAF (N-terminal)
		5J7T	UCH and UBL1-3
		7VIJ	UBL1-5 (C-terminal)

2.1.12 Databases

Table 23: Databases used

Database	Application	Link
AlphaFold Protein Structure Database	Protein structure prediction and retrieval	https://alphafold.ebi.ac.uk
AlphaMissense	Prediction of mutation pathogenicity	https://alphamissense.hegelab.org
Basic Local Alignment Search Tool (BLAST)	Nucleotide sequence alignment	https://blast.ncbi.nlm.nih.gov
cBioPortal for Cancer Genomics	Mutation analysis and visualization	https://cbioportal.org/

National Center for Biotechnology Information (NCBI)	General access to biological databases and tools	https://ncbi.nlm.nih.gov/
NCBI’s Gene Expression Omnibus (GEO)	Obtaining publicly available genomics data	https://ncbi.nlm.nih.gov/geo/
NCBI’s Primer-BLAST	Primer designing tool	https://ncbi.nlm.nih.gov/tools/primer-blast
NEB Tm Calculator	Annealing temperature calculation	https://tmcalculator.neb.com/
The Cancer Dependency Map Project at Broad Institute (DepMap)	Cancer cell line dependency analysis	https://depmap.org
The Cancer Therapeutics Response Portal	Drug screen data analysis	https://portals.broadinstitute.org/ctrp.v2.1/
The Human Protein Atlas	Lymph node image obtention	https://proteinatlas.org/
The Protein Data Bank (PDB)	Protein 3D structure database	https://rcsb.org
Uniprot	Protein database	https://uniprot.org/

2.1.13 Software

Table 24: Software used

Software	Application	Developer	Link
Affinity Designer (version 1.10.0)	Vector graphics editor	Serif, Nottingham, United Kingdom	https://affinity.serif.com
AlphaFold	Protein structure prediction	DeepMind, London, United Kingdom	https://alphafold.ebi.ac.uk
BioRender	Illustration and cartoon creation	BioRender, Toronto, Canada	https://biorender.com
ChemoStarTS	Immunoblotting imaging	SCPA, Milan, Italy	https://scpa.it
FlowJo (version 10.10.0)	Flow cytometry analysis	BD Biosciences, Franklin Lakes, New Jersey, USA	https://flowjo.com
Gen5 (version 3.04)	Microplate reader data acquisition	BioTek Instruments (Agilent Technologies), Winooski, Vermont, USA	https://biotek.com/gen5
GraphPad Prism (version 10.2.3)	Graphs and statistics	GraphPad software Inc., San Diego, California, USA	http://graphpad.com
Integrative Genomics Viewer (IGV) (version 2.16.2)	Genomics data visualizer	University of California, San Diego, La Jolla, California, USA; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA	https://igv.org
Microsoft Office 365	Microsoft Office 365	Microsoft Corporation, Redmond, Washington, USA	https://www.office.com
Morpheus	Data visualization	Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA	https://software.broadinstitute.org/morpheus

Quantum CX5 Edge (version 18.06)	Gel documentation	Vilber Lourmat, Collégien, France	https://vilber.com
Robetta	Protein structure prediction	Baker Lab, University of Washington, Seattle, Washington, USA	https://robetta.bakerlab.org
RStudio (version 2021.09.0+351)	Graphs and RNA-seq analysis	Posit PBC, Boston, Massachusetts, USA	https://posit.co/
SnapGene Viewer (version 7.0.2)	Primer design and DNA sequence visualization	SnapGene, Chicago, Illinois, USA	https://snapgene.com
StepOne Software (version 2.3)	Conducting and analyzing quantitative real-time PCR	Applied Biosystems (Thermo Fisher Scientific), Waltham, Massachusetts, USA	https://thermofisher.com/stepone
The PyMOL Molecular Graphics System (version 2.5.8)	Molecule or protein viewer and renderer	Schrödinger, Inc., New York City, New York, USA	https://pymol.org
ZEN (version 2.3)	Confocal image acquisition	Carl Zeiss Microscopy, Oberkochen, Germany	https://zeiss.com

2.2 Methods

2.2.1 Molecular biology techniques

2.2.1.1 Polymerase chain reaction

Polymerase chain reaction (PCR) was performed to amplify DNA sequences using gene-specific primers. A reaction mixture with a total volume of 25 µL was prepared, containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.4 µM of each forward and reverse primer, 0.5 units of Phusion High-Fidelity DNA Polymerase, and 50-100 ng of template DNA. The PCR was carried out in a thermal cycler using the cycling conditions summarized in **Table 25**. PCR products were visualized on a 0.8% agarose gel stained with Midori green and imaged.

Table 25: Thermal cycling conditions for polymerase chain reaction

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	98	2 minutes	1
Denaturation	98	30 seconds	30
Annealing	Primer-specific	30 seconds	
Extension	72	30 seconds/kb	
Final extension	72	5 minutes	1

2.2.1.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments based on their size. DNA samples were loaded onto a 0.8% agarose gel stained with Midori green, and an electric field was applied. DNA migration was visualized under UV light, and fragment sizes were compared against a molecular weight marker.

2.2.1.3 DNA extraction from agarose gel

DNA fragments were excised from agarose gels following electrophoresis and purified using the Wizard® SV Gel and PCR Clean-Up kit, following the manufacturer's instructions. The extracted DNA was quantified using Nanodrop, used immediately or stored at -20°C until further use.

2.2.1.4 Gateway recombination cloning

To investigate gene function, the recombination-based Gateway® cloning system was utilized to generate expression constructs. The ORF of USP7 was amplified either entirely or partially, in the case of truncated versions, and attB sites were added using specific primers (***Table 15***). PCR products were run in a 0.8% agarose gel and the expected nucleotide band size was extracted from the gel and cloned into an attP sites-containing pDONR™221 entry vector through BP recombination reaction using Gateway® BP Clonase™ according to the manufacturer's instructions. The product of the BP reaction, containing attL sites (***Figure 8***), was transformed into chemically competent *E. coli* by incubating the mixture on ice for 30 min, followed by heat shock for 45 seconds at 42°C and subsequent recovery by incubation under agitation for one hour at 37°C. The transformed bacteria were plated on LB agar plates containing kanamycin and incubated at 37°C for a maximum of 16 h. The next day, individual bacterial clones were picked and expanded in LB medium supplemented with 50 µg/mL kanamycin and incubated for 12-16 h at 37°C. Plasmids were purified using the Qiagen miniprep kit, following the manufacturer's instructions. The resulting clones were verified by sequencing using M13 primers.

2.2.1.5 LR recombination reaction

The different USP7 variants were flipped from the entry clone into the destination vectors pLX307 or pLX312 (***Figure 8***) via LR recombination reaction using LR Clonase. Similarly, the attL site-containing pCR8/GW/TOPO PCGF1 plasmid was cloned into the pLX307 destination vector via LR recombination reaction. The expression clones were transformed into chemically competent bacteria as previously described. Transformed bacteria were plated onto ampicillin LB agar plates and incubated at 37°C for a maximum of 16 h. The following day, bacterial clones were selected and expanded in LB medium supplemented with 100 µg/mL ampicillin and incubated for 12-16 h at 37°C. Plasmids were purified using the Qiagen miniprep kit, following the manufacturer's instructions. All expression constructs were validated by sequencing using EF1alpha primers.

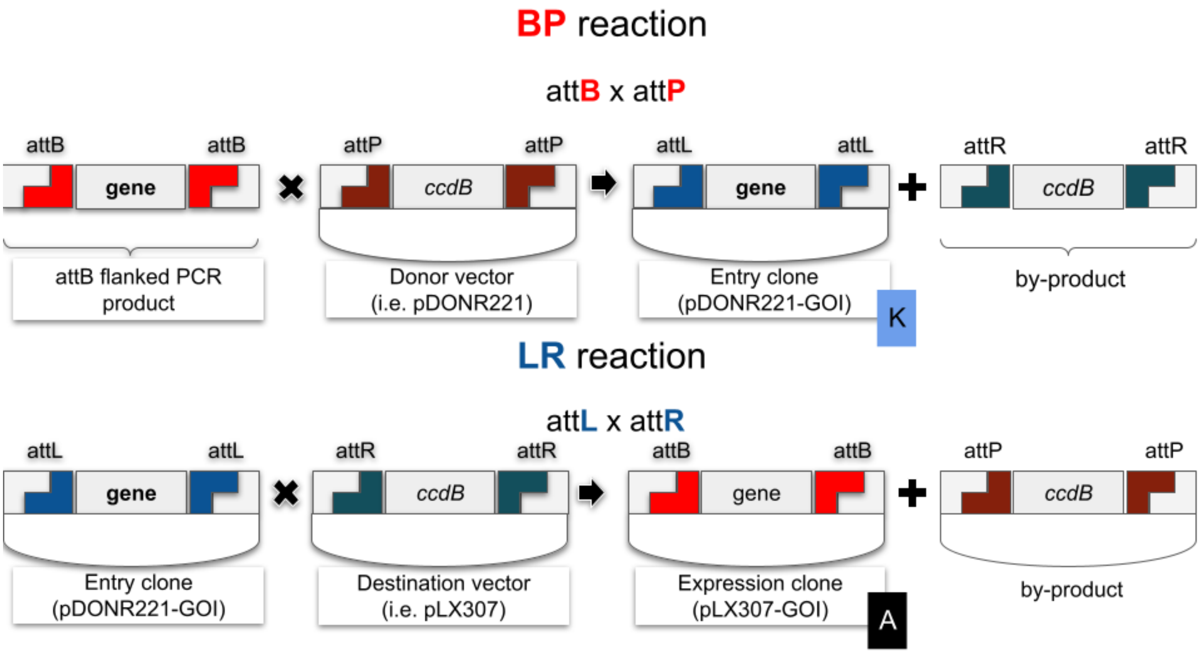

Figure 8: Schematic representation of the Gateway™ BP and LR recombination reactions. In the BP reaction, the gene of interest (GOI) flanked by attB sites is recombined into the donor vector (pDONR221) containing attP sites, generating the entry clone (pDONR221-GOI). In the LR reaction, the entry clone is recombined with a destination vector (pLX307) containing attR sites, producing an expression clone (pLX307-GOI). The ccdB gene is excised as a by-product in both reactions. K, kanamycin; A, ampicillin. Schematic cartoon generated by D. Joppi.

2.2.1.6 Site-directed mutagenesis

Site-directed mutagenesis was performed using PCR with specific primers for USP7 mutations (***Table 19***) using 100ng of pDONR221-USP7 as a template. For the PCR, the reaction mixture consisted of 12.5 µL of NEBNext® High-Fidelity 2X PCR Master Mix, the template, 10 µM of both forward and reverse primers, and nuclease-free water to a final volume of 25 µL. The PCR products were run on a 0.8% agarose gel, and the bands were excised as previously described. The product was then phosphorylated with T4 Polynucleotide Kinase (PNK) for two hours, followed by overnight ligation at 16°C. The next day, the reaction product was transformed into chemically competent bacteria and kanamycin-resistant clones were purified with Qiagen miniprep kit, as previously described. LR reaction was performed as previously described and the USP7 variants were cloned into the destination vector pLX312. The same procedure adopted for LR reaction products as previously described was conducted and expression constructs were validated by sequencing using specific primers designed for *USP7* for point mutation verification (***Table 16***).

2.2.1.7 Purification of plasmid DNA

To isolate plasmid DNA, mini- and midiprep kits were used. Minipreps were always the preferred choice, except for expression plasmids. For minipreps, kanamycin- or ampicillin-resistant bacterial colonies were picked and cultured overnight in 4 mL of LB broth containing either 50 µg/mL kanamycin or 100 µg/ml ampicillin. Plasmid DNA was then purified following the manufacturer's instructions. For expression plasmid preparations, which require higher concentrations of endotoxin-free plasmid DNA for transfection, midipreps were performed using 200 mL of overnight bacterial culture added with the respective antibiotic. The isolation of plasmid DNA was then conducted using NucleoBond® Xtra Midi EF Kit according to the manufacturer's instructions.

2.2.1.8 RNA extraction and cDNA synthesis

For RNA extraction, 5 x 10^6 cells were harvested and washed with cold PBS to remove any residual media or contaminants. The washed cell pellet was resuspended in 1 mL of TRIzol™ reagent and the next steps were followed according to the manufacturer's instructions. The concentration and purity of the extracted RNA were determined using a Nanodrop spectrophotometer, ensuring that the 260/280 ratio was within the optimal range of 1.8-2.0. For the synthesis of complementary DNA (cDNA), 500 ng of total RNA was used as the **t**emplate, and the RevertAid® first strand synthesis kit with OligodT primers was used according to the manufacturer's instructions.

2.2.1.9 Real-time quantitative polymerase chain reaction

Real-time quantitative PCR (RT-qPCR) was performed using the GoTaq® qPCR Master Mix following the recommendations from the manufacturer. The qPCR reactions were set up in a 96-well qPCR plate, with each reaction prepared at least in triplicates. The plate was briefly centrifuged to ensure that all components were at the bottom of the wells and to remove any air bubbles that might interfere with fluorescence detection. The qPCR was run using qPCR cycler and fluorescence data were collected at each cycle to monitor the amplification of the target gene in real-time. The resulting cycle threshold (Ct) values were used to quantify relative gene expression levels, normalized to the expression of the housekeeping gene GAPDH. The ΔΔCt method was employed for relative quantification, comparing the gene expression levels in treated *versus* control samples.

2.2.2 Cell biology assays

2.2.2.1 Cell culture

Cell lines were cultured with their appropriate media, as detailed in **Table 6**, and were split twice a week for maintenance. Suspension cells (DLBCL cell lines) were split by dilution with fresh medium to maintain optimal density. Adherent cells (HEK29ET) were washed with PBS, detached using 0.05% trypsin-EDTA, resuspended, and seeded at appropriate densities. All cell lines were maintained at 37°C in a humidified incubator containing 5% CO_2 to ensure optimal growth conditions.

2.2.2.2 Transfection and lentiviral transduction

To produce lentiviral particles, 2 x 10^6 HEK293T cells were seeded into 10-cm tissue culture dishes in DMEM supplemented with 10% FBS and incubated overnight at 37°C in a humidified atmosphere containing 5% CO_2 to allow the cells to adhere and reach the optimal confluency of 60-80% by the time of transfection. Cells were transfected using either PEI or TransIT-293 transfection reagent. Either PEI or TransIT were mixed with 1 ml Opti-MEM and incubated at RT for 5 min. 10 µg of the desired transfer plasmid was used for each transfection, along with 7.5 µg of the packaging plasmid psPAX2 and 5 µg of the envelope plasmid pMD2.G. For PEI-mediated transfection, the DNA and PEI were mixed at a 1:3 ratio (µg:µg) and incubated at room temperature (RT) for 30 min to allow the formation of DNA-PEI complexes. For TransIT-mediated transfection, the DNA was mixed with the TransIT reagent at a 1:2 ratio (µg:µL) and incubated for 30 min at RT. The transfection mixture was then added dropwise to the HEK293T cells, and the cells were returned to the incubator to allow DNA uptake. Transfection using TransIT was performed using D30 medium, while PEI-mediated transfection was done using D10, which was changed to D30 after 6 h. Viral supernatants were harvested at 24 and 48 hours (h) post-transfection. Medium containing the virus particles was filtered through a 0.45 µm filter to remove any cells. The filtered virus supernatants were either used immediately for transduction or stored at -80 °C for later use. DLBCL cells were prepared by resuspending them in their appropriate cell medium at a density of 2 x 10^6 cells per mL. The viral supernatant was added to the target cells in the presence of 8 µg/mL of polybrene. The cells were then centrifuged at 1,000 x g for 90 min at 30°C. After centrifugation, the cells were incubated at 37°C in 5% CO_2 overnight, and the medium was replaced with fresh culture medium the following day. Three days post-transduction, selection of successfully transduced cells was initiated by the addition of 2 µg/mL of puromycin to the culture medium. The antibiotic selection was maintained for 3 days, during which time the death of non-puromycin resistant (non-transduced) cells was induced, and only cells harboring the pLX307 or lentiGuide-puro plasmids, which confer puromycin resistance, survived. The selected, stably transduced cells were expanded and used for downstream experiments, ensuring that the transduction was efficient and that the desired gene expression or deletion was achieved. As for cells transduced with pLX312 constructs, which include GFP as a selection marker, sorting was performed one week post-transduction. Before sorting, cells were washed with PBS and filtered using a 70 µm strainer to prevent clumping and ensure a single-cell suspension. Cell sorting was conducted using a BD FACSAria flow cytometer. The sorted cells were then further maintained under standard culture conditions for expansion and downstream assays.

2.2.2.3 Proliferation assay

To assess cell proliferation, 5-20 x 10^3 cells were seeded in quadruplicate into each well of a 96-well plate in a final volume of 100µL per well, 1 plate per day of measurement. The cells were cultured at 37°C in a humidified incubator with 5% CO_2 to maintain optimal growth conditions. Cell proliferation was monitored for five days. Every 24h, 10µL of Alamar Blue reagent was added directly to each well containing the cultured cells. After the addition of Alamar Blue, the plates were incubated for 3h at 37°C in a humidified incubator with 5% CO_2. Fluorescence was measured using a spectrophotometer, with the excitation wavelength set between 530-560nm and the emission wavelength measured at 590nm.

2.2.2.4 Drug assay

To investigate the effect of the USP7 inhibitor FT671 in DLBCL cell lines, 20 x 10^3 cells were seeded in quadruplicates in a 96-well plate in a final volume of 100µL per well, added with the drug. The concentrations of FT671 used ranged from 48.8nM to 50µM, using DMSO as the vehicle (control). Cells were incubated for 72h at 37°C in a humidified incubator with 5% CO_2. On the final day, cell viability was assessed by adding Alamar Blue and incubating for 3 hours, as described previously.

2.2.2.5 Determination of the half-maximal effective concentration

To determine the half-maximal effective concentration (EC50) of FT671, nonlinear regression analysis was implemented. Dose-response curves were generated from the viability assay readout with FT671, and EC50 values were calculated using GraphPad Prism (version 10.2.3). EC50 values were defined as the concentration of FT671 required to achieve 50% of the maximal effect in each experimental condition.

2.2.2.6 Immunofluorescence and imaging

For immunofluorescence staining, $2x10^4$ cells were prepared either by centrifugation using a Cytospin at 200 rpm for 5 min or by seeding them onto poly-D-lysine-coated glass coverslips in 6-well plates and allowing them to adhere overnight at 37°C in a humidified atmosphere with 5% CO_2. After attachment to the glass coverslips or slides, cells were fixed with 4% PFA in PBS for 10 min at RT to preserve cellular structures. Following fixation, the slides were washed once with 1x PBS to remove excess PFA. Permeabilization of the cells was performed using 0.1% Triton X-100 in PBS for 5 min and washed three times in 1x PBS. Samples were subsequently blocked for 1h at RT with 5% BSA in PBS. After blocking, the slides were incubated overnight at 4°C in a humidified chamber with the specific primary antibody, which was diluted in PBS containing 0.5% BSA. The following day, slides were washed three times with 1x PBS and incubated with secondary antibodies, specific to the species of the primary antibody, conjugated with Alexa Fluor 488, Alexa Fluor 546 or Alexa Fluor 647 fluorophores. The secondary antibody was diluted in PBS with 0.5% BSA, and the slides were incubated for 1h at RT in the dark to protect the fluorophores from photobleaching. After the secondary antibody incubation, the slides were washed three times with 1x PBS to remove unbound secondary antibodies. For nuclear staining, a solution of 1µg/ml of DAPI was added to the slides, which were then incubated for 5 min at RT. Following DAPI staining, slides were washed once more with 1x PBS and mounted in mounting medium to preserve the samples and minimize refractive index differences during imaging. Coverslips were carefully placed on top of the slides to avoid air bubbles. Images were acquired with the LSM780 Zeiss confocal laser scanning microscope with a 67x oil objective.

2.2.3 Protein biochemistry

2.2.3.1 Cell lysis and protein extraction

Cells were harvested by centrifugation at 400 x g for 5 min at 4°C and washed twice with cold DPBS to remove any residual culture media and non-cellular debris. The cell pellet was resuspended in ice-cold RIPA lysis buffer supplemented with a protease and phosphatase inhibitor cocktail to prevent protein degradation and dephosphorylation. Cells were lysed by incubating the suspension on ice for 30 min, with occasional vortexing to ensure thorough lysis. Following lysis, the cell debris was removed by centrifugation at 20,000 x g for 15 min at 4°C. The supernatant containing the total protein lysate was collected. For subcellular fractionation, 5×10^6 to 1×10^7 were collected, centrifuged at 400 x g for 5 min at 4°C and washed twice with cold PBS. The cell pellet was resuspended in 1x Cao lysis buffer for lysis and incubated on ice for 30 min, with occasional mixing. Samples were centrifuged for 5 min at 1,000 x g at 4°C and the supernatant containing the cytosolic fraction was collected. The nuclei was washed twice in cold Cao lysis buffer and the pellet was resuspended in 1x RIPA lysis buffer for nuclear membrane disruption. Samples were incubated for 30 min in ice, with occasional vortexing. Samples were centrifuged at for 15 min at 20,000 x g at 4°C and the supernatant containing the nuclear fraction was collected and frozen at -20°C until usage. Following lysate preparation, the protein concentration was determined using the Lowry-based DC assay, following the manufacturer's protocol.

2.2.3.2 Immunoprecipitation

For immunoprecipitation, 10µL of Pierce protein G-coated magnetic beads were used per reaction. The beads were washed three times with RIPA lysis buffer and incubated with 3 µg of the specific primary capture antibody or its corresponding isotype control in a final volume of 500 µL of 1x RIPA buffer overnight at 4°C on a rotator to ensure efficient antibody binding. Following capturing, the antibody-bound beads were washed three times with cold 1x RIPA buffer using a magnetic separator to remove any unbound antibodies. For HA tag immunoprecipitation, magnetic beads pre-coated with an anti-HA antibody were utilized. These beads were also washed three times in 1x RIPA buffer to remove any preservatives or contaminants. Subsequently, 500µg of total protein lysate, was added to the capture antibody-bound Protein G magnetic beads or anti-HA magnetic beads. The mixture was incubated overnight at 4°C with rotation to allow for the specific binding of target proteins to the antibody-coated beads. Following this incubation, the beads were washed four times with cold RIPA buffer to remove nonspecifically bound proteins. After washing, the bound proteins were eluted from the beads by adding LDS sample buffer containing reducing agent to ensure proper protein elution. Additionally, the samples were boiled at 95°C for 10 min to disrupt any protein-protein interactions and to fully denature the proteins, thereby ultimately releasing them from the magnetic beads. The magnetic beads were then separated from the eluted proteins using a magnetic separator, and the supernatant containing the eluted proteins was collected. The eluted proteins were subsequently subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting to analyze the presence of the target proteins and their potential interacting partners.

2.2.3.3 Protein separation and Immunoblotting

For SDS-PAGE, 10µg of total protein from each sample were mixed with LDS Sample Buffer (4x) and Reducing Agent (10x) and denatured by boiling at 95°C for 10 min. The denatured protein samples were loaded onto Bis-Tris based 8% acrylamide gels and run using MES SDS running buffer for 15 min at 80V, followed by 1h and 15 min at 120V. Size-separated proteins were transferred from the gel onto a PVDF membrane for 2 h at 20V using transfer buffer to ensure efficient protein transfer from the gel to the membrane. Following the transfer, the PVDF membrane was blocked for 1h at RT with gentle agitation with TBS/T and 5% BSA to prevent nonspecific antibody binding.

After blocking, the membrane was incubated overnight at 4°C with the desired primary antibody diluted in TBS/T containing 0.5% BSA. The next day, membranes were washed three times with TBS/T to remove any unbound primary antibody, followed by incubation with a species-specific HRP-conjugated secondary antibody diluted in TBS/T for 1h at RT. The membrane was washed again in TBS/T before detecting the signal using ECL substrate. Protein bands were visualized using a chemiluminescent imaging system.

2.2.3.4 Membrane stripping and reprobing

To detect additional proteins from the same PVDF membrane, membranes were stripped using either sodium azide or Restore™ stripping buffer, depending on the species origin or the isotype of the antibody used. For antibodies from different species and isotypes, the membrane was incubated for 2h with 0.2% sodium azide under agitation at RT in TBS/T, followed by washes with TBS/T. The membrane was then probed overnight with the respective primary antibody. When antibodies from the same species and isotypes were used, however, Restore™ stripping buffer was applied. Membranes were incubated at RT for 10-15 min, followed by thorough washing in TBS/T to ensure the removal of any residual stripping solution. Membranes were then re-blocked with 5% BSA in TBS/T for 1h before being probed again with a new primary antibody.

2.2.4 Computational analysis

2.2.4.1 Protein structure and protein-protein interaction prediction

In collaboration with Amal Hassan Ali from the Institute of Medical Physics and Biophysics in Charité, Berlin, *in silico* structural predictions of PRC1.1 protein complex members, BCOR, USP7, BCORL1, and PCGF1 have been performed. The predictions of the human wild-type proteins were based on their amino acid sequence. Initial trials were done with DeepMind AlphaFold. However, due to the size of some proteins, it was not possible to predict their full-length structures using a single method. Therefore, Robetta was implemented for the predictions, and, in combination with available structures from the PDB, the full-length PRC1.1 complex was successfully modeled. For all the analysis, we used PyMOL and Chimera X-1.3. In addition to the wild-type USP7, several mutations observed in DLBCL patients—Y243D, I660K, T730S, and Y1056H—were generated *in silico* using PyMOL.

2.2.4.2 Single protein structure prediction

For parts of the proteins that have not been crystallized to the present moment, predictions of parts or the protein 3D structure were conducted using Robetta and AlphaFold software. Several structures of the USP7 were available and used, such as the PDB IDs, 4M5X and 5J7T, to have a quality control by superimposing the structures with our predicted model of USP7. The PDB ID 5KYB represents the 3D configuration of the active site of USP7. Whereas the PDB ID 5J7T was aligned and superimposed with the C-terminal of the protein. The full-length BCOR structure predicted from the AlphaFold was only achievable by dividing the amino acid sequence into three segments (See supplements): (1) Starting from the N-terminal 1-600 aa, (2) 601-1185aa, and (3) 1186 aa to the C-terminal. Most of the protein segments were predicted as loops and short alpha-helices, indicating a lack of defined structure.

2.2.4.3 Protein-protein interaction prediction

To facilitate the detection of interactions between PRC1.1 complex members, we made use of already crystallized structures: PCGF1/BCOR interaction (PDB ID: 4HPL) and PCGF1/BCOR/KDM2B (PDB ID: 8HCU) interactions. In addition, we generated predictions of the interaction between BCOR and USP7 using Robetta and AlphaFold. The following FASTA sequences of USP7 and BCOR were used for the contact predictions (***Table 27***).

Table 26: USP7 and BCOR FASTA sequences used for preliminary contact predictions in Robetta

BCOR FASTA sequence (amino acids: 1636 to 1755)
VFEFEFSETPLLPCYNIQVSVAQGPRNWLLLSDVLKKLKMSSRIFRCNFPNVEIVTIAEAEF YRQVSASLLFSCSKDLEAFNPESKELLDLVEFTNEIQTLLGSSVEWLHPSDLASDNYW
USP7 FASTA sequence (amino acids: 181 to 300)
KGFIDDDKVTFEVFVQADAPHGVAWDSKKHTGYVGLKNQGATCYMNSLLQTLFFTNQLR KAVYMMPTEGDDSSKSVPLALQRVFYELQHSDKPVGTKKLTKSFGWETLDSFMQHDVQ ELC

2.2.5 Statistical analysis

Fisher's exact test was implemented to assess whether mutations in PRC1.1 subunits were enriched in a specific genetic or transcriptional subtype of DLBCL and if they co-occurred or exhibited mutual exclusivity. The unclassified cases from the ABC and GCB-type analysis included both the unclassified cases and those that did not have this classification information available (na). A separate contingency table analysis was performed for each gene, and p-values were calculated, with a significance threshold set at 0.05. For the co-occurrence analysis, pairwise comparisons of mutations in *BCOR*, *USP7*, *BCORL1*, *KDM2B,* and *EZH2* were performed, and p-values were calculated. followed by q-value calculation to adjust for multiple comparisons using the false discovery rate (FDR) method. A q-value threshold of 0.1 was considered significant.

Additionally, t-tests were performed to compare means between two groups where appropriate. The t-test was applied to assess whether the groups had statistically significant differences for continuous variables. For all tests, p-values less than 0.05 were considered statistically significant. Statistics were conducted using GraphPad Prism (version 10.2.3).

3 Results

3.1 Frequent mutations in Polycomb group proteins in Diffuse Large B-cell Lymphoma

Polycomb group proteins (PcGs) are key epigenetic modifiers that regulate gene expression with frequent genetic alteration and subsequent oncogenic dysregulation in cancers. Given the central role of *EZH2* mutations for the pathophysiology of C3 DLBCL and the sparsity of data for other PRCs' potential involvement in DLBCL pathogenesis, I aimed to investigate whether PRC1 subunits are commonly mutated in DLBCL and how these mutations might contribute to lymphomagenesis and/or disease progression.

To investigate the role of PRCs, I analyzed two large whole-exome sequencing studies comprising 878 DLBCL patients[25,26]. This analysis revealed that members of the PRC1.1 complex were mutated our cohort in 12% (37/304) of DLBCL patients and 10% (56/574) of DLBCL patients in the Schmitz cohort resulting in an overall frequency of in 11% (93/878) of DLBCL patients (***Figure 9A*** and ***Table 27***), suggesting a potential role for PRC1.1 dysfunction in the pathogenesis of DLBCL. The PRC1.1 complex subunits frequently mutated are *USP7*, *BCOR*, *BCORL1,* and *KDM2B* (***Figure 9B*** and ***C***), prompting our decision to focus on PRC1.1 for subsequent functional studies on PRC1.1. Other PRC1 complexes, such as PRC1.2-5, had less frequent mutations, and only the PRC1.6 complex reached similar frequencies as PRC1.1. Notably, the mutations are predominantly non-synonymous (***Table 27***), suggesting an impact on protein functionality and potential activity disruption of the complex in DLBCL.

Table 27: Mutation count and characteristics of Polycomb repressive complexes in DLBCL.

PRC subcomplex	Chapuy *et al.*[25]	Schmitz *et al.*[26]	Total	Non-synonymous (%)	Synonymous (%)
PRC1.1	37	56	93	75.3	24.7
PRC1.6	25	57	82	79.3	20.7
PRC1.3/5	23	45	68	88.2	11.8
PRC1.2/4	12	24	36	60.5	39.5
PRC2	33	39	72	86.1	13.9

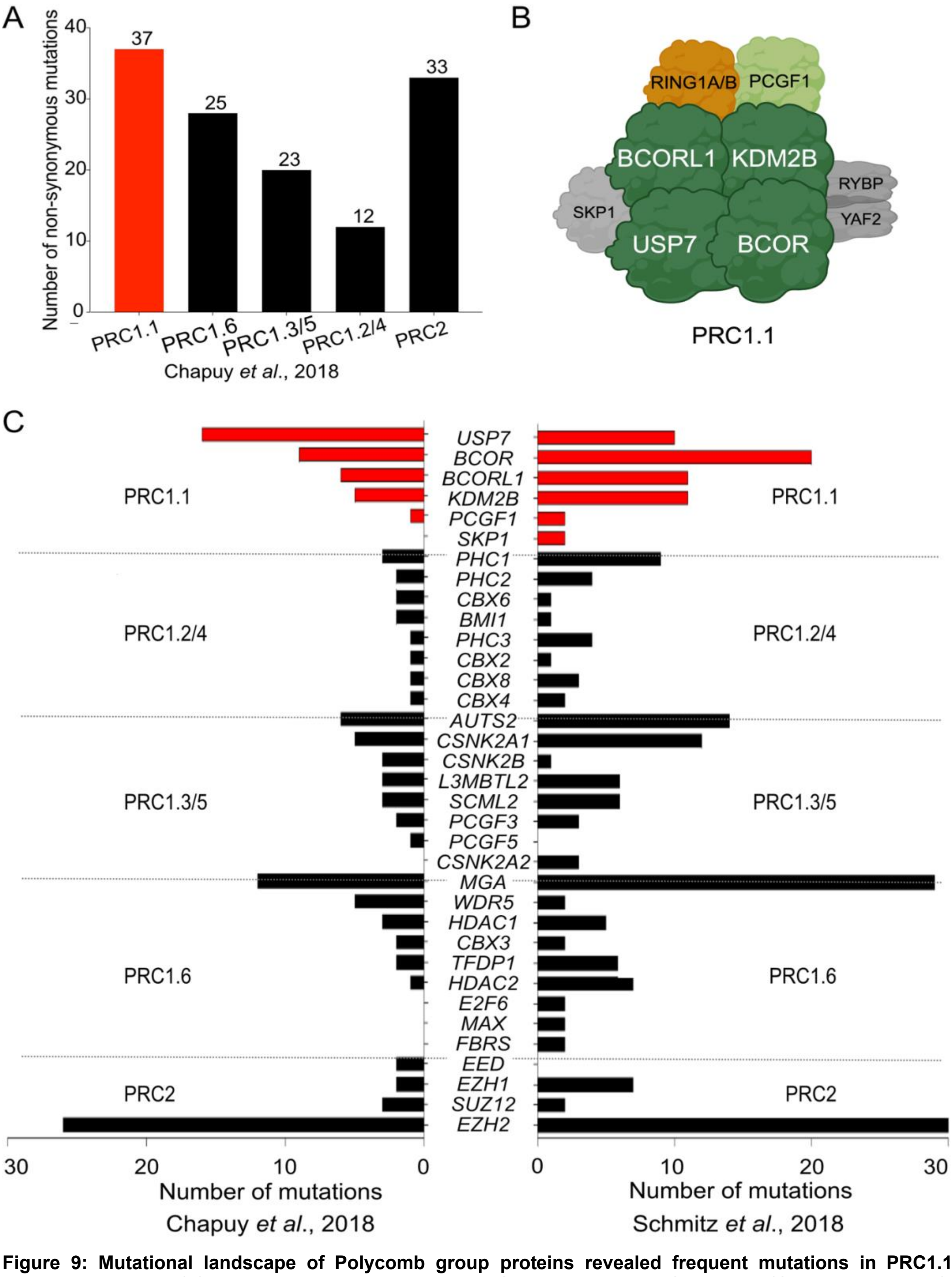

Figure 9: Mutational landscape of Polycomb group proteins revealed frequent mutations in PRC1.1 complex members. (A) Non-synonymous mutations in PRC1 members in DLBCL patients (Chapuy *et al*., 2018). (B) Cartoon representation of PRC1.1 including its core components (PCGF1 and RING1A/B) and subunits (USP7, BCOR, KDM2B, BCORL1, SKP1, RYBP and YAF2). (C) Distribution of mutations of PRC1.1 subunits (red) and other Polycomb group proteins in two independent DLBCL patient cohorts (Chapuy *et al*., 2018, to the left and Schmitz *et al*., 2018, to the right).

3.2 Characterization of PRC1.1 mutations in DLBCL subtypes

3.2.1 PRC1.1 mutations are predominantly missense and distributed in DLBCL

To gain a better understanding of frequent PRC1.1 mutations in DLBCL and their impact, I analyzed the characteristics of the mutations that occur in *BCOR, USP7, BCORL1, KDM2B, PCGF1,* and *SKP1*. The majority of the non-synonymous mutations in PRC1.1 subunits were identified as missense mutations (***Table 28*** and ***Figure 10A***). *BCOR* harbored up to 18% truncating mutations, while *USP7* had up to 25% mutations affecting splice sites. In general, PRC1.1 mutations appear to occur heterogeneously across patients (***Figure 10A***), suggesting the need for closer examination or further stratification to determine if there is a particular subgroup where these alterations are enriched.

Table 28: Mutation type distribution of PRC1.1 members in DLBCL patients (n= 878):

Gene	Total number	Missense		Nonsense		Frameshift		Splice site		In frame	
		Total	%	Total	%	Total	%	Total	%	Total	%
BCOR	22	14	63.6	4	18.2	3	13.6	1	4.55	0	0
USP7	19	13	68.4	1	5.26	1	5.26	4	25.1	0	0
BCORL1	13	11	84.6	1	7.69	1	7.69	0	0	0	0
KDM2B	9	9	100	0	0	0	0	0	0	0	0
PCGF1	3	2	66.7	0	0	0	0	0	0	1	33.3
SKP1	1	1	100	0	0	0	0	0	0	0	0

3.2.2 *USP7* mutations are enriched in GCB-type DLBCL

First, I grouped patients based on the COO classification (GCB and ABC), which also includes unclassified cases (Uncl.), to detect a possible enrichment of non-synonymous mutations within these groups (***Table 29*** and ***Figure 10B***). To identify a putative co-occurrence of the PRC1 mutations with PRC2, *EZH2* mutations were included, as *EZH2* mutations are known to be enriched in GCB and C3-type DLBCL[25,26,41]. *USP7* mutations were notably enriched in GCB-type DLBCL, like *EZH2* mutations typically are, with 11 mutations observed in the GCB-type, compared to 3 mutations in ABC-type DLBCLs and 5 in the unclassified DLBCL cases. Fisher's exact test confirmed that the enrichment of *USP7* mutations in GCB-type DLBCL was statistically significant ($p = 0.0049$). In contrast, no statistically significant difference in mutation frequencies was observed for *BCOR*, *BCORL1*, or *KDM2B* between GCB and ABC subtypes (p-values equal to 0.584, 0.882, and 0.731, respectively). These findings suggest a potential subtype-specific role for USP7 in GCB-type DLBCL.

Table 29: Mutations in PRC1.1 members across COO DLBCL subtypes

COO subtype		Cases	*BCOR*		*USP7*		*BCORL1*		*KDM2B*	
			Mut	WT	Mut	WT	Mut	WT	Mut	WT
ABC		408	12	396	3	405	3	405	4	404
GCB		258	5	253	11	247	3	255	4	245
Uncl.		212	5	207	5	207	6	206	0	212
p-value	ABC *vs.* GCB		0.584		0.0049		0.882		0.731	
	ABC *vs.* Uncl.		0.871		0.186		0.0863		0.359	
	GCB *vs.* Uncl.		1.0		0.380		0.330		0.177	

3.2.3 PRC1.1 mutations are mutually exclusive and do not co-occur with *EZH2* mutations

To understand whether each of the mutations in PRC1.1 subunits significantly overlap within the same patients or genetic subtypes or COO groups or are mutually exclusive, I performed an analysis of co-occurrence and mutual exclusivity (***Figure 10C***). Specifically, mutations in *USP7* and *EZH2* did not co-occur, as evidenced by the lowest q-value of 0.11, which was not statistically significant. Similarly, no other mutations were co-occurring statistically (***Figure 10C***). Given the clear absence of co-occurrence between PRC1.1 and *EZH2* mutations, further analyses focused exclusively on PRC1.1 subunits, excluding *EZH2* from subsequent assessments.

3.2.4 *USP7* mutations are frequent in C3 and C4 clusters

When assessing the PRC1.1 members in the genetic subtypes (C1-C5 DLBCLs) that concentrate a high number of GCB-type DLBCL cases, namely C3- and C4-DLBCL (n = 207), I found that *USP7* mutations were significantly enriched (p-value < 0.0001) in these DLBCL subtypes compared to other DLBCL clusters (***Figure 10D***). *BCOR, BCORL1, KDM2B* and *PCGF1* did not exhibit a significant enrichment in C4/C4 DLBCLs (p-value > 0.05). The enrichment of *USP7* mutations in C3 and C4 DLBCLs suggests that these mutations may play a crucial role in the pathogenesis of this subtype.

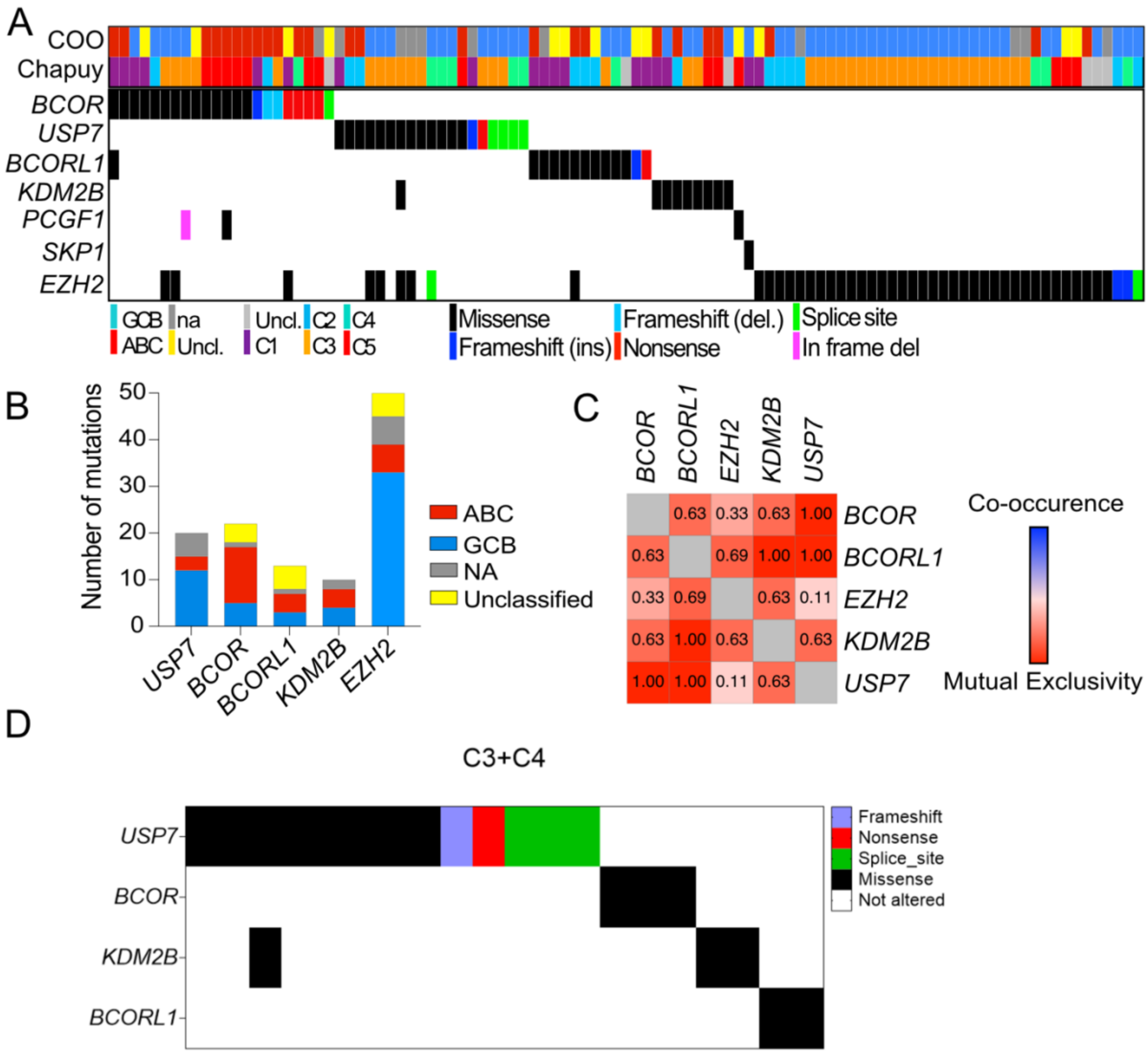

Figure 10: Mutation landscape and co-occurrence of PRC1.1 subunits in DLBCL. Patient data were analyzed from Chapuy *et al.* and Schmitz *et al.* (n=878 primary DLBCL). (A) Mutation distribution of PRC1.1 subunits (*BCOR, USP7, BCORL1, KDM2B, PCGF1* and *SKP1*) as well as *EZH2* is shown across DLBCL patients, stratified by cell-of-origin classifications (GCB, blue; ABC, red; unclassified, yellow) and molecular clusters (C1, purple; C2, blue; C3, orange; C4, cyan; C5, red; unclassified, grey). Mutations are classified by type: missense, black; frameshift insertion, dark blue; frameshift deletion, light blue; nonsense, red; splice site, green; in-frame deletions, pink. *USP7* and *BCOR* exhibit high mutation frequencies, particularly in GCB-like and C3/C4 clusters. (B) Number of mutations in *USP7, BCOR, BCORL1, KDM2B,* and *EZH2* in ABC, GCB, and unclassified DLBCL subtypes. (C) Heatmap depicting q-values representing mutual exclusivity of mutations between *BCOR, BCORL1, EZH2, KDM2B,* and *USP7*. (D) Mutational distribution of mutation types in the PRC1.1 subunits *USP7, BCOR, KDM2B,* and *BCORL1* in DLBCL patients stratified by C3 and C4 molecular subtypes. Uncl., unclassified; na, not available; ins., frameshift insertion; del., frameshift deletion.

3.3 Distribution of Recurrent Mutations in PRC1.1 Components

Next, I examined the distribution of PRC1.1 complex member mutations across the coding sequences (CDS) and the known functional domains of *USP7*, *BCOR*, and *BCORL1.* The mutations in *USP7*, including those specifically occurring in C3 and C4 clusters (***Figure 11***), highlighted in orange, are spread across the CDS without a hotspot and this pattern is often found in genes that exhibit a tumor suppressor function. Most mutations in *USP7* are missense (green circles), although two truncating mutations (black circles) are also observed, particularly in important domains of the gene i.e. ubiquitin C-terminal hydrolase (UCH) and the C-terminal ubiquitin-like (UBL) domains, respectively. The UCH domain is the catalytic core of USP7, responsible for its deubiquitinating activity, and UBL domains have been described to be crucial for the full activity of USP7[102,103]. Hence, mutations in *USP7* could disrupt its deubiquitinase activity, leading to dysregulation of USP7 target proteins, such as RING1B[104] and p53[103], which could potentially contribute to lymphomagenesis through altered regulation of chromatin modification and apoptosis.

BCOR mutations are, similar to those seen in *USP7*, widely distributed across the CDS with a few truncating mutations and missense mutations located across the CDS of the gene (***Figure 11***), again indicative of a putative tumor suppressor function. Notably, mutations in *BCOR* seem to cluster around the N-terminal, as well as in the ankyrin repeat region (ANK) and PCGF Ub-like fold discriminator (PUFD) domain regions. ANK and PUFD are one of the few structured regions of BCOR. While the function of the ANK domain is not fully known, the PUFD domain is particularly important for interaction with PCGF1[99], further suggesting that these alterations may impact BCOR's structural integrity as well as functional activity as part of PRC1.1 in DLBCL.

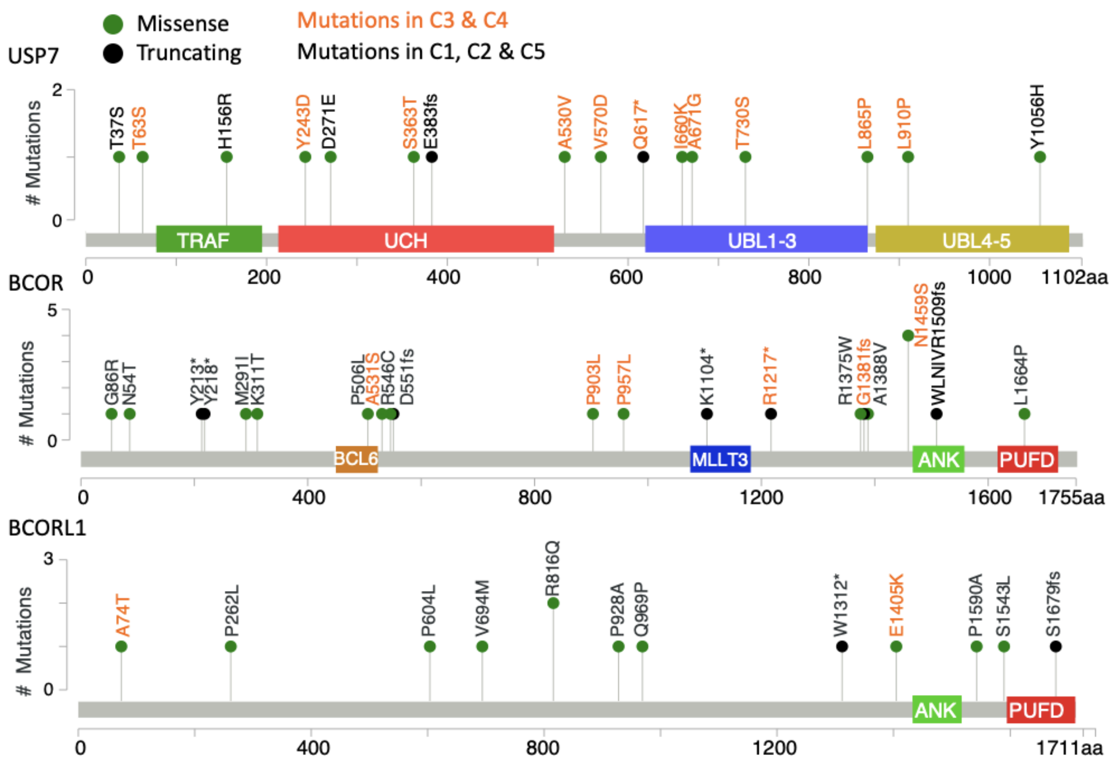

Figure 11: Mutational pattern of PRC1.1 subunits from DLBCL patients. Mutations in *USP7, BCOR,* and *BCOLR1* are depicted. Missense mutations are shown with green circles, and truncating mutations have black circles. Mutations occurring in C3 and C4 clusters are highlighted in orange, and mutations in the remaining clusters (C1, C2, and C5) are labeled in black. Lollipop figures were generated with the mutation mapper at cBioPortal.

Similarly, for *BCORL1*, mutations are distributed across the gene (***Figure 11***), with both missense mutations and truncations occurring close to the ANK and PUFD domains, which can also suggest an impairment of interaction with PCGF1 in DLBCL through its PUFD domain[87], like BCOR.

3.4 USP7 and BCOR are upregulated in GCB-type DLBCL

Having assessed the alterations in PRC1.1 members at the DNA level, I next focused on their expression patterns in DLBCL patient data across distinct subtypes. I queried the expression

of USP7 in the DLBCL patient cohort from Chapuy et al. across distinctly classified groups. Looking at patients from a COO classification perspective. This analysis revealed that the expression of *USP7* and *BCOR* was significantly higher in the GCB subtype ($p = 0.0057$ and 0.00238, respectively). Conversely, *PCGF1* did not show any statistically significant difference across the group. (***Figure 12A***). Similarly, *KDM2B* did not have an enrichment of expression (data not shown) in any group. Further stratification of the DLBCL patient data into the C1-C5 molecular subtypes revealed that *USP7* is significantly upregulated in the C3 and C4 clusters compared to C1, C2, and C5 ($p = 0.0142$) (***Figure 12B*** and ***C***). In contrast, *BCOR* and *PCGF1* did not show significant expression differences across the C3 and C4 clusters versus the other clusters (***Figure 12B*** and ***C***).

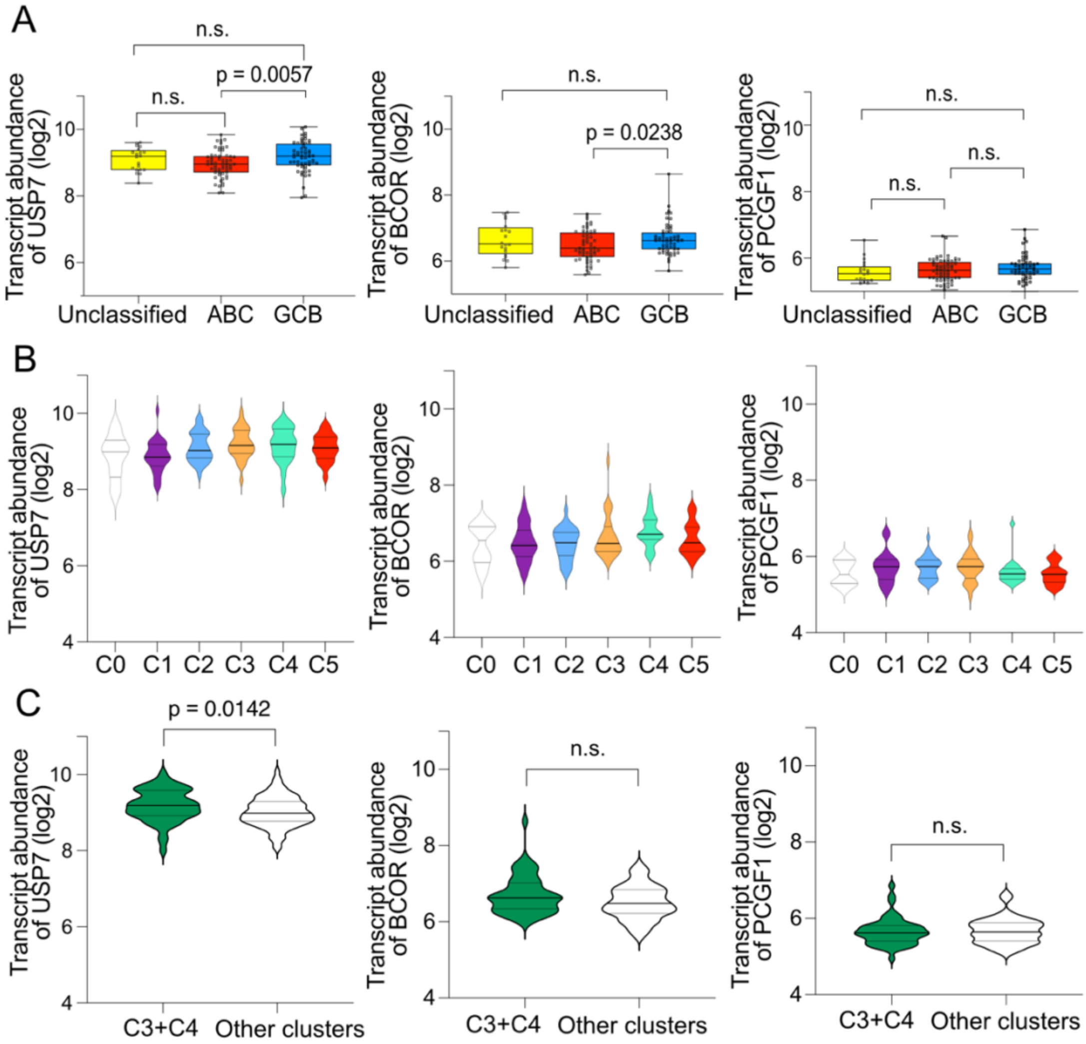

Figure 12: Expression of USP7, BCOR and PCGF1 in DLBCL. Expression of USP7, BCOR and PCGF1 was assessed in the Chapuy et al. dataset across COO subtypes (A) or molecular subtype classification (B) and (C). (A) Box plots showing transcript abundance (log2) of USP7, BCOR, and PCGF1 in distinct COO subtypes: Unclassified (n = 20), ABC (n = 61), and GCB (n = 56). USP7 and BCOR are significantly upregulated in the GCB subtype compared to ABC. PCGF1 shows no statistical difference in expression in none of the COO subtypes ($p > 0.05$). n.s., non-significant. (B) Violin plots displaying transcript abundance (log2) of USP7, BCOR, and PCGF1 in five molecular subtypes (C1–C5, C0 being unclassified). (C) Violin plots comparing transcriptional abundance (log2) of USP7, BCOR, and PCGF1 in the combined C3 and C4 clusters versus other clusters (C1, C2 and C5). A t-test was implemented to compare two groups. n.s., non-significant.

3.5 PRC1.1 subunits are upregulated in germinal center B-cells.

By leveraging publicly available data[105], I queried the expression of PRC1.1 subunits during normal B-cell development (***Figure 13A***). Once naïve B-cells are activated and differentiate into GC B cells (union of centroblasts and centrocytes), PRC1.1 subunits are upregulated, as has been previously observed by Beguelin and colleagues[59]. Notably, the expression of *USP7* reached levels comparable to key regulators of the GC reaction, such as *EZH2* and *BCL6*, a finding that had not been previously appreciated. Similarly, *KDM2B* and BCOR display significantly high transcript levels in GC B cells, compared to naïve B cells. PCGF1 and BCORL1, on the other hand, do not show a dramatic increase in the GC reaction. IHC staining from the lymph node of a female individual confirmed high levels of USP7 specifically in the GC (***Figure 13B***), comparable to the GC marker BCL6. This suggests that USP7 may play a critical role in GC biology and may have a previously underexplored role in the development of DLBCL, making it an interesting candidate for further investigation. Conversely, PCGF1 expression was not particularly elevated in the GC (***Figure 13B***), as it is also expressed in regions extra-GC in the lymph node.

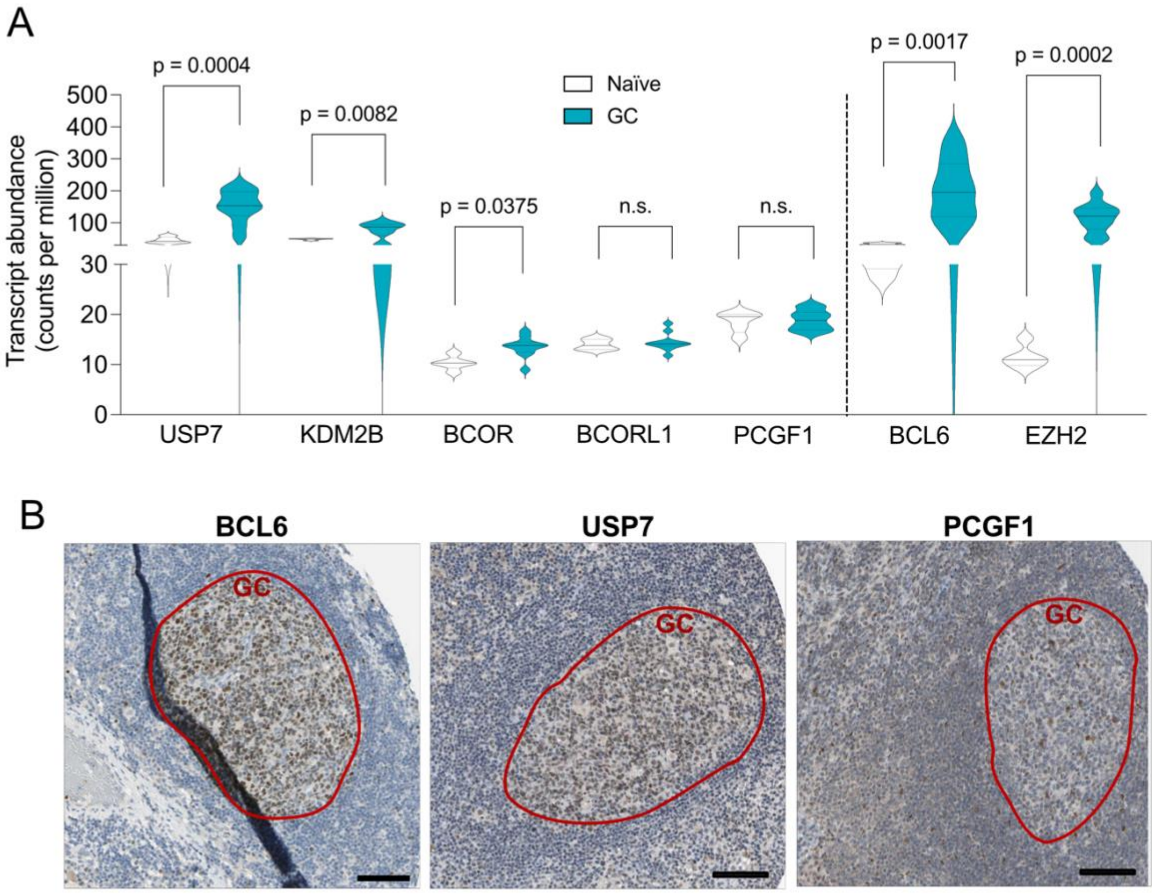

Figure 13: Expression of PRC1.1 subunits across normal B cell development. (A) Transcript levels of PRC1.1 subunits, EZH2, and BCL6 in naïve B cells (naïve) and germinal center B cells (GC). T-test was employed as statistical analysis to compare two groups. (B) Immunohistochemistry of BCL6 (positive control for the germinal center region), USP7, and PCGF1 in lymph node samples from 53 (female), 55 (male), and 53-year-old (female) individuals, respectively (images obtained from The Human Protein Atlas). The germinal center (GC) compartment is circled in red. Black scale bar corresponds to 100 µm.

3.6 Characterization and selection of model cell systems

Having assessed the mutation patterns in primary DLBCL samples, I next focused on exploring *in vitro* models to better understand the molecular mechanisms underlying PRC1.1 dysregulation and finding the most suitable model system for this study. Specifically, I analyzed WES- and RNA-sequencing data generated by our laboratory from 29 DLBCL cell lines (Table 6) to evaluate whether the mutation frequency observed in patient samples was also reflected in these models.

3.6.1 Assessment of mutations of PRC1.1 complex members in model cell lines

I explored the expression and mutation patterns of PRC1.1 subunits across these DLBCL cell lines, first by assessing whether the high number of mutations in PRC1.1 seen in DLBCL was also reflected in the cell lines. The data was filtered for non-synonymous mutations in PcGs. Out of the 29 cell lines analyzed, 18 harbored mutations for any of the PcGs. I observed a similar pattern as seen in patients, with a total of 16 non-synonymous mutations in PRC1.1 subunits in DLBCL cell lines, the highest number seen among the PRCs, including PRC2 (***Figure 15A***), emphasizing the relevance of PRC1.1 components in DLBCL. Mutations in PRC1.1 occurred mainly in *BCOR*, *USP7*, *BCORL1* and *KDM2B*, predominantly in GCB-type DLBCL cell lines (***Figure 15B***).

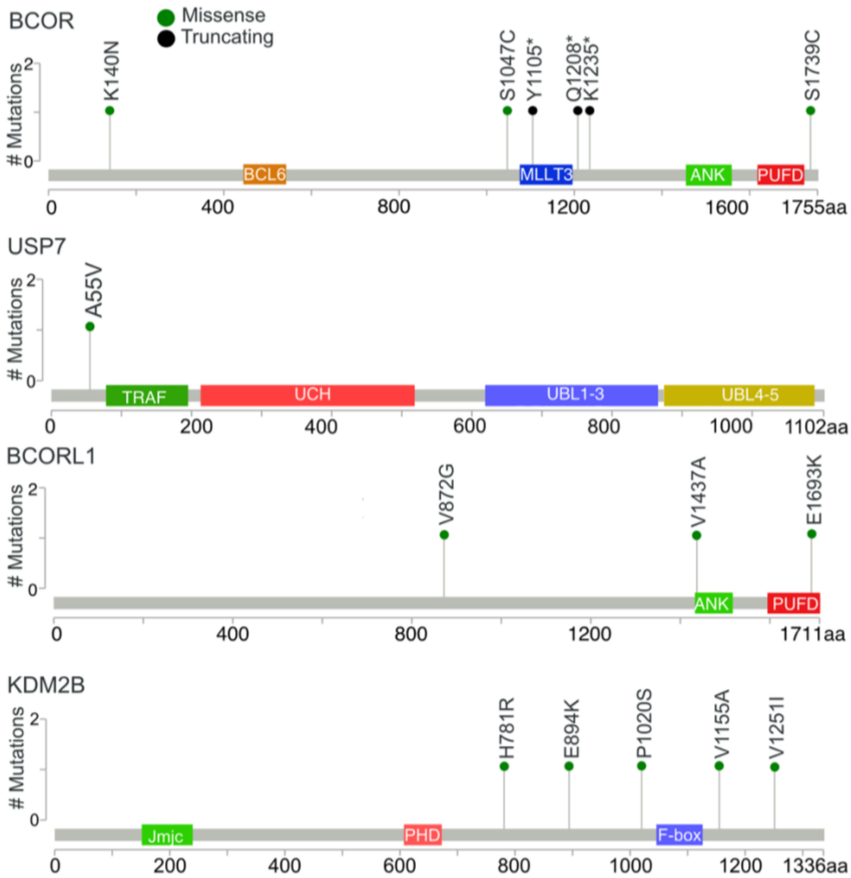

Figure 14: Mutational pattern of recurrent PRC1.1 mutations in *USP7*, *BCOR*, *BCORL1,* and *KDM2B* in DLBCL cell lines. Mutations in *USP7*, *BCOR*, *BCORL1*, and *KDM2B* are depicted. Missense mutations are highlighted as green circles, and truncating mutations are shown as black circles. Lollipop figures were generated with mutation mapper at cBioPortal and modified by D. Joppi.

3.6.2 Assessment of transcript and protein abundance of PRC1.1 complex members in model cell lines

Next, I analyzed the transcript abundance of USP7, BCOR, and PCGF1 across the 29 DLBCL cell lines (***Figure 15C***). USP7 expression was higher compared to BCOR and PCGF1 in most cell lines. PCGF1 displayed consistently low transcript levels in many cell lines, while BCOR had a co-expression with USP7 in certain cell lines such as LY3, U2932, K422, DHL7, LY7, Toledo, TMD8, DHL8, A4/FUK, DHL4, Pfeiffer, among others. In terms of protein levels, BCOR and USP7 were indeed co-expressed in some cell lines, such as the ABC-type DLBCL cell lines LY3, HBL1, U2932, DHL2, and in GCB-type DHL8, Toledo, Pfeiffer, and DHL4 (***Figure 15D***).

3.7 Assessment of PCGF1 expression and function in DLBCL model cell lines

3.7.1 Levels of PCGF1 are low in DLBCL cell lines

The PCGF1-6 proteins are the core component defining the non-canonical PRC complexes. To investigate the specific role of PRC1.1 in DLBCL, I assessed the expression of PCGF1 in the DLBCL cell lines, comparing it to other PCGF proteins (PCGF2–6), which make up the remaining PRC1 complexes (***Figure 16A***). In accordance with primary DLBCL data, PCGF1 expression is significantly low in DLBCL cell lines compared to other PCGFs, particularly PCGF3 and PCGF5 ($p<0.0001$ and $p<0.0001$, respectively). Similarly, I found low transcript abundance PCGF2, PCGF4, and PCGF6 (***Figure 16A***).

These findings prompted three important and interconnected questions: (1) Why are PCGF1 levels low in DLBCL? (2) Is the PRC1.1 complex formed in DLBCL cell lines? (3) What is the functional consequence of PCGF1 overexpression in DLBCL?

3.7.2 Ectopic expressed PCGF1 co-localizes with USP7 in the nucleus

To understand this, I transduced an informative DLBCL cell line panel, comprised of ABC- and GCB-type DLBCL cell lines, with pLX307-PCGF1-V5 and successfully generated cell lines stably overexpressing PCGF1 (***Figure 16C***). Notably, enforced expression of PCGF1 resulted in increased expression of the H2AK119Ub mark, highlighting the functionality of the assembled complex (***Figure 16B*** and ***C***).

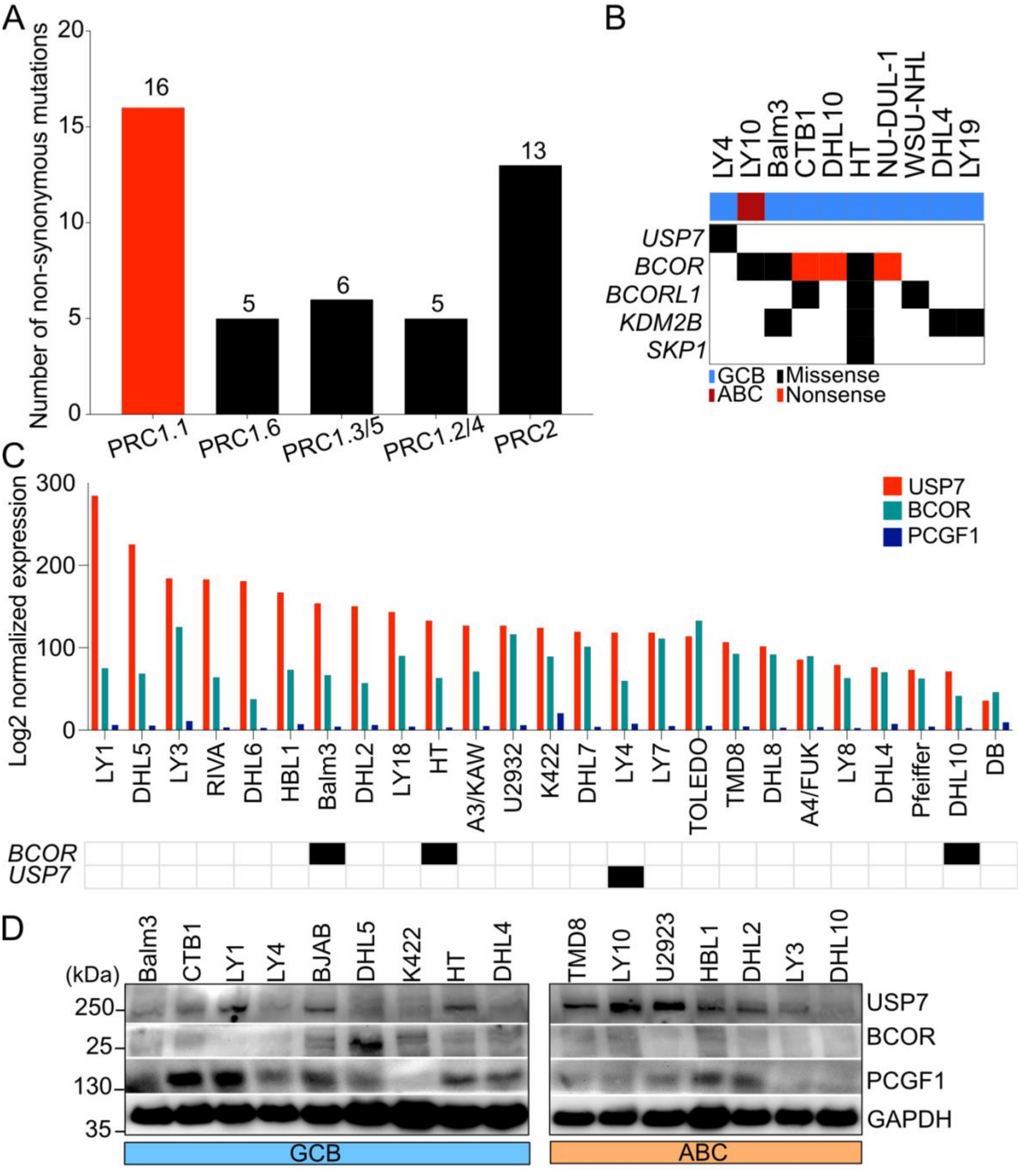

Figure 15: Characterization of PRC1.1 in DLBCL cell lines. (A) High number of non-synonymous mutations in PRC1.1 subunits, compared to other Polycomb repressive complexed, observed in DLBCL cell lines. (B) Heatmap depicting missense and nonsense mutations in PRC1.1 members (*USP7*, *BCOR*, *BCORL1*, *KDM2B* and *SKP1*) in DLBCL cell lines stratified by their cell-of-origin: activated B-cell-type (ABC) or germinal center B-cell-type (GCB). (C) RNA levels of *USP7*, *BCOR* and *PCGF1* in a panel of 25 DLBCL cell lines. Heatmap underneath the graph depicts mutations for BCOR or USP7 (black, mutated; WT, white). (D) Protein levels of USP7, BCOR and PCGF1 in a panel of GCB- and ABC-type DLBCL cell lines. GAPDH was used as loading control.

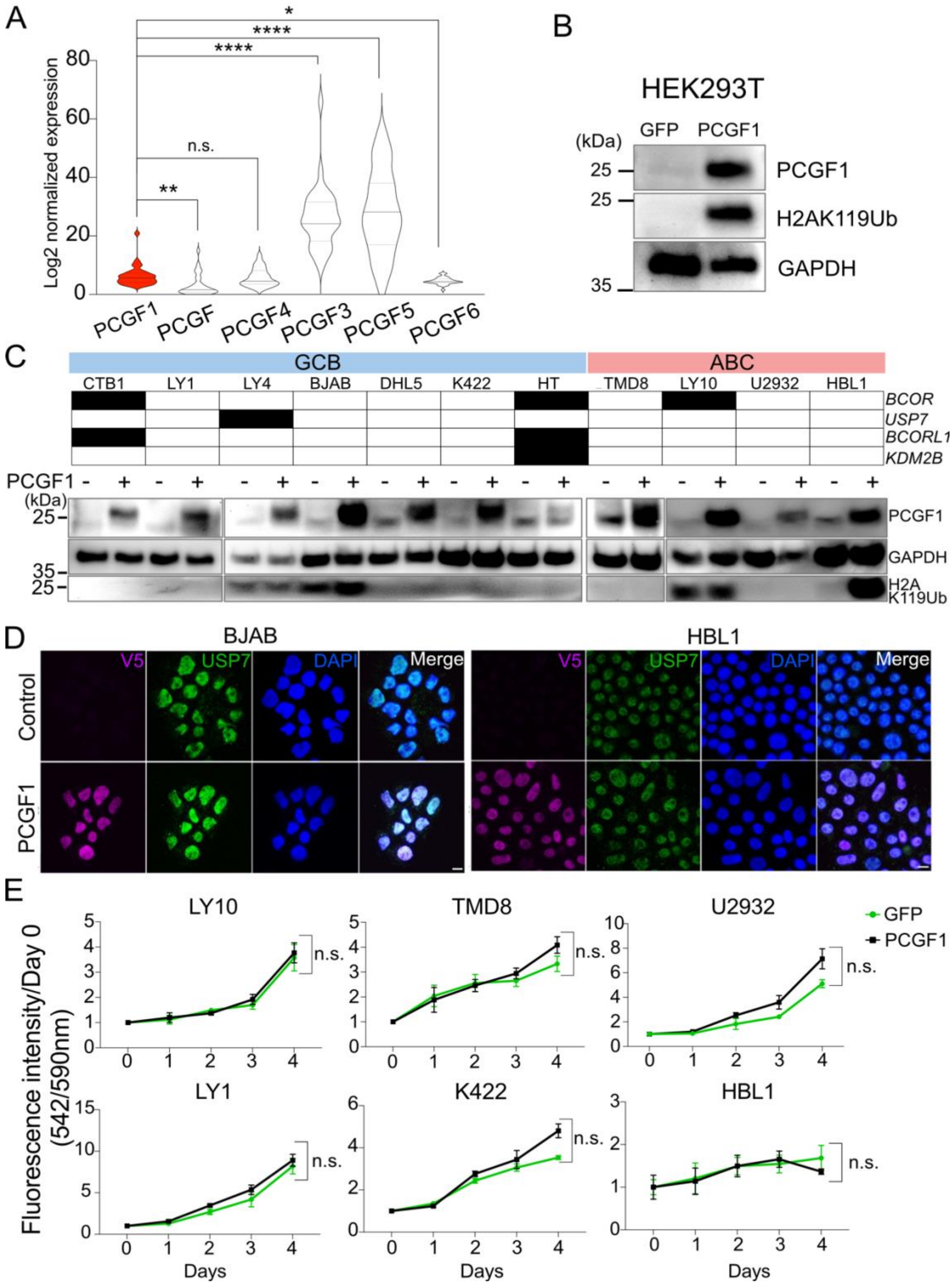

Figure 16: Ectopic PCGF1 has nuclear localization and increases H2AK119Ub. (A) Violin plots display relative log expression levels of PCGF1 to PCGF6 in DLBCL cell lines. PCGF1 shows the lowest expression, while PCGF3 and PCGF5 exhibit higher relative expression. (B) Enforced expression of PCGF1 or GFP control in HEK293T cells using a pLX307 vector. Immunoblot for indicated proteins confirms ectopic expression of PCGF1. (C) Confirmation of PCGF1 overexpression in a panel of GCB and ABC-type DLBCL cell lines shows changes in H2AK119Ub mark. GAPDH was used as loading control. Mutational status for *BCOR*, *USP7*, *BCORL1,* and *KDM2B* for the cell lines panel is indicated in the upper panel (black color denotes mutation; white represents no mutation). (D) Ectopic PCGF1 co-localized with USP7 in the nucleus. Staining against V5-tagged PCGF1 and USP7 was performed. DAPI was used as nuclear marker. White bar corresponds to 10µm. (E) Proliferation of PCGF1 vs. GFP-overexpressing cells assessed with alamarBlue across 4 days in LY10, TMD8, U2932, LY1, K422, and HBL1.

To further investigate the connection between PCGF1 and other PRC1.1 components, I evaluated the intracellular localization of PCGF1 and endogenous USP7 in HBL1 and BJAB cells using immunofluorescence (***Figure 16D***). Imaging revealed that both exogenous PCGF1 and endogenous USP7 were predominantly localized to the nucleus, prompting us to focus next on the role of USP7. Following their growth, using GFP-overexpressing cells transduced from the same backbone as a control, no significant effect on cellular proliferation was observed in the cells with enforced expression of PCGF1 (***Figure 16E***).

3.8 Characterizing the role of USP7 in DLBCL

3.8.1 Genetic knockout of USP7 does not affect cell viability

To investigate the importance of USP7 for DLBCL, I designed three single guide RNAs (sgRNAs) specifically targeting the *USP7* gene in exons 8, 17, and 15 to generate genetic ablation of *USP7* in DLBCL cell lines.

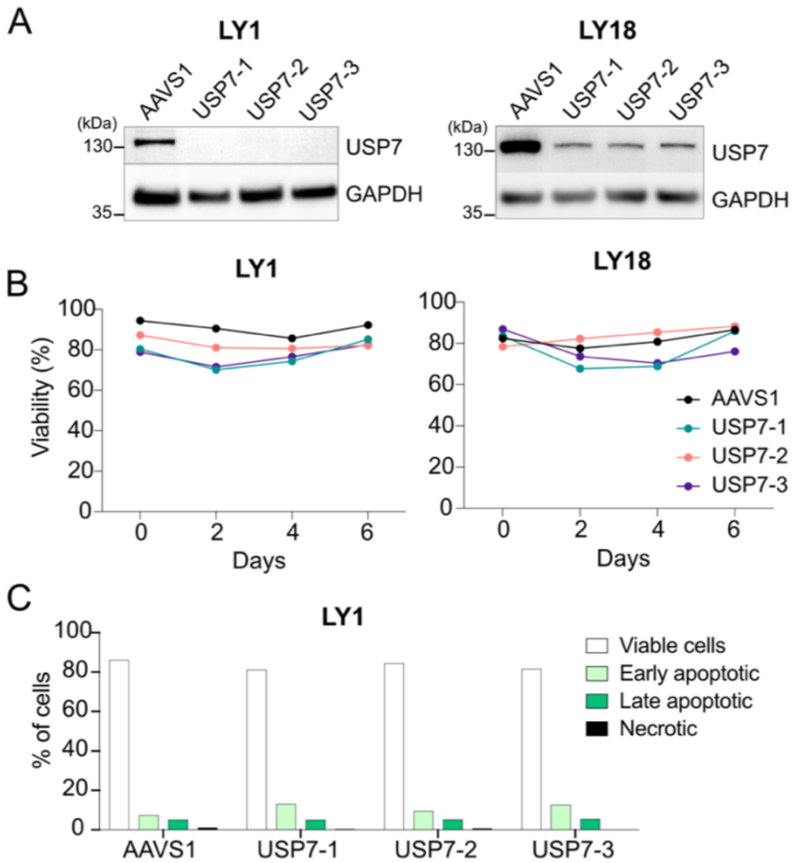

Figure 17: Genetic depletion of USP7 in GCB-type DLBCL cell lines. (A) Confirmation of Cas9-mediated USP7 knockout in LY1 and LY18 cell lines. GAPDH was used as loading control. (B) Viability assessment of GCB DLBCL cell lines LY1 and LY18 across 6 days using propidium iodide shows that these cell lines are not dependent on USP7. The ratio between live and dead cells (unstained/stained, respectively) was used to measure viability. (C) Annexin V/PI staining of LY1 cells on day 5 after puromycin selection shows that USP7 KO cells do not show any effect on cell viability when compared to the AAVS1 control.

These sgRNAs were used for screening in GCB DLBCL cell lines given its increased expression in GCB primary and cell lines. Following this approach, an efficient knockout (KO) of USP7 in two GCB cell lines, LY1 and LY18, was achieved and documented using western blot against USP7 (***Figure 17A***). The gRNAs targeting the AAVS1 locus, a non-transcribed locus, was used as control. Next, I checked whether *USP7* depletion altered cell proliferation of the resulting LY1 and LY18 cell lines by measuring the living cells using flow cytometry and propidium iodide (PI) to stain dead cells. No viability difference of the cells with the three USP7 gRNAs were detected compared to the control sgRNA (AAVS1) (***Figure 17B***), indicating that these DLBCL cell lines are not dependent on USP7 to survive. In addition, another GCB DLBCL cell line, LY1 cells, which in contrast to the above two GCB lines also harbors oncogenic gain-of-function *EZH2* mutations were stained with Annexin V and PI to detect viable, early and late apoptotic cells. Again, no effect on cell death was seen upon USP7 genetic depletion in these cells (***Figure 17C***).

3.8.2 DLBCL cells are resistant to USP7 inhibition

Besides its structural role within the PRC1.1 complex, USP7 (also referred to as "HAUSP") also acts as a deubiquitinase. To this end, I asked if the enzymatic function of USP7 affect cell survival or proliferation using a pharmacological inhibitor against USP7. Due to its function as an oncogene and representation as a potential therapeutic target in various diseases such as AML, there are several specific inhibitors available [85]. Treatment with the USP7 inhibitor FT671, which binds irreversibly to the catalytic domain of USP7, was performed in a panel of DLBCL cell lines for 72h using a range of concentrations up to 50µM. The calculated concentration for FT671 to induce cell death or cell cycle arrest in at least 50% of the cells (EC50) was high for all cell lines tested (***Figure 18***). Similar results have been seen with another USP7 inhibitor (HBX-41108) used in DLBCL lines[106–108] (***Table 30***):

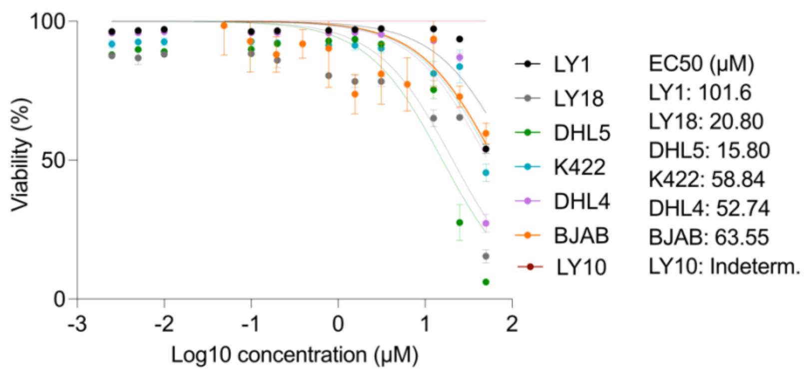

Figure 18: Effect of the pharmacological inhibition of USP7 on viability of DLBCL cell lines treated with FT671. Proliferation of DLBCL cell lines upon 72-hour treatment with USP7 inhibitor (FT671) is affected only upon high drug concentration exposition (right panel indicates the EC50). Viability was normalized to cells treated with vehicle (DMSO) only. Indeterm., indeterminable.

Table 30: EC50 of HBX-41108 in DLBCL cell lines.

Cell line	Cell-of-origin	EC50 (µM)
LY10	ABC	>100
LY19	GCB	23.10
DHL8	GCB	23.02
K422	GCB	19.12
Toledo	GCB	19.96
LY3	ABC	13.53
DHL4	GCB	10.71
DB	GCB	5.93
NUDHL1	ABC	6.14
PFEIFFER	ABC	3.52

3.8.3 USP7 inhibition reduces monoubiquitination of histone H2A at lysine 119

To investigate whether the inhibition of USP7 would be sufficient to reverse the H2AK119Ub mark in parental cells, I treated DLBCL cell lines with 3 µM of FT671 for 48 h. I found that H2AK119Ub was decreased upon FT671 treatment in cell lines, such as K422 and DHL2 (***Figure 19***), suggesting that the inhibition or the impairment in USP7 deubiquitinase activity might be enough to disturb the PRC1.1 complex in certain DLBCL cells. However, despite using effective concentration, I did not observe a reduction of the H2AK119Ub mark in LY7 parental cells. One interesting observation is the reduction of USP7 protein abundance following the inhibition of the USP7 enzymatic activity. This data suggests that the USP7 enzymatic activity can modulate the H2AK119Ub mark and prompted the following studies assessing the interplay of PCGF1 and USP7 for the H2AK119Ub mark abundance.

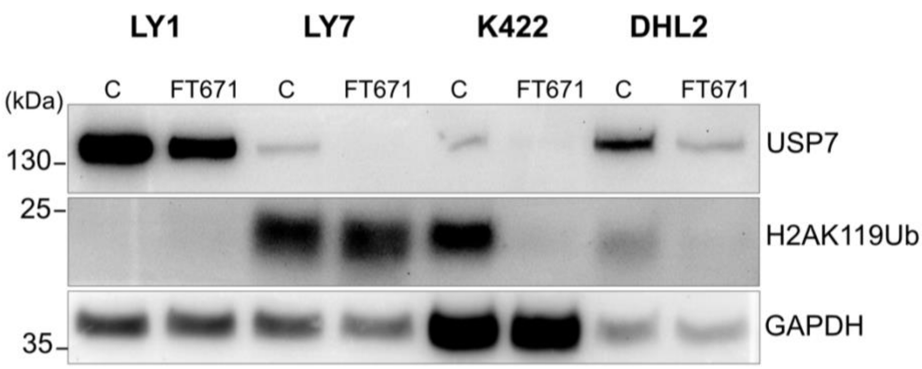

Figure 19: Pharmacological inhibition of USP7 induces H2AK119Ub changes. Immunoblotting depicting DLBCL cell lines treated with 3 µM for 48h leads to a decrease in PRC1 histone mark in K422 and DHL2 cells. GAPDH was used as loading control.

To assess if and to which extent the PCGF1-induced changes in H2AK119Ub are reversed by USP7 inhibition, the DLBCL cell lines HBL1, K422, and LY1 overexpressing PCGF1 were subjected to USP7 inhibition using the USP7 inhibitor FT671. All tested cell lines exhibit a strong induction of the H2AK119ub mark following the enforced expression of PCGF1 (***Figure***

20A and ***B***). Interestingly, all cell lines also have a strong downregulation of the H2AK119ub mark following USP7 inhibition, pronounced in the GCB cell lines (***Figure 20B***) compared to the ABC-cell line HBL1 (***Figure 20A***).

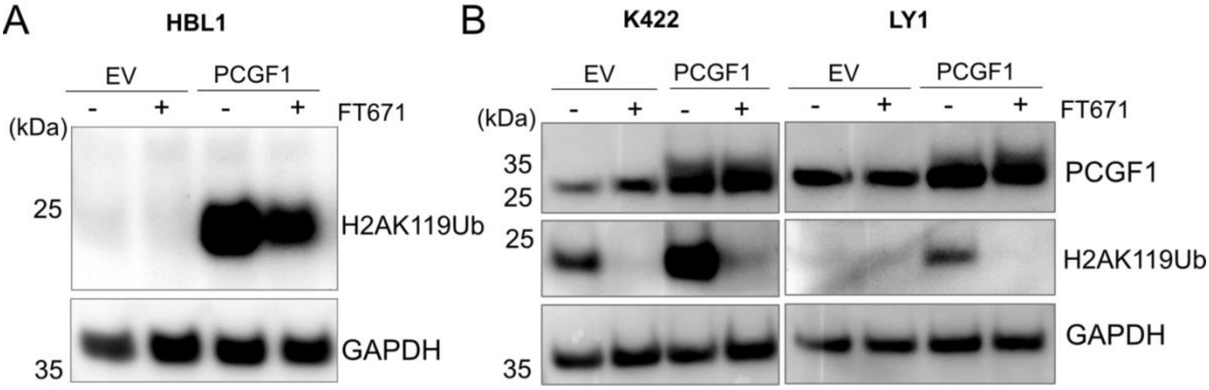

Figure 20: Inhibition of USP7 in PCGF1 overexpressing cell lines leads to a reduction in PCGF1-induced H2AK119Ub mark. (A) HBL1, (B) K422 and LY1 PCGF1-overexpressing cells show an increased H2AK119Ub mark compared to GFP-overexpressing cells (control) and this effect is reduced partially in HBL1 treated with 3 µm FT671 for 48h. GAPDH was used as loading control.

3.9 Full-length USP7 interacts with BCOR and ectopic PCGF1

To investigate if direct interactions between USP7 and PCGF1 occur in DLBCL cell lines, immunoprecipitation of endogenous USP7 was performed in PCGF1-V5-overexpressing cells. The immunoprecipitates were probed against V5 to detect putative binding to the ectopically expressed PCGF1 and endogenous BCOR. The results indicate an interaction of USP7 in the ABC cell model HBL1 and the GCB cell model BJAB between exogenous PCGF1 and USP7 (***Figure 21A***). Furthermore, USP7 was also found to interact with endogenous BCOR, and the protein did not interact with EZH2 in these two cell lines. These findings underscore the specificity of the interaction between PRC1.1 members and endogenous USP7, providing important insights into the molecular dynamics governing their association in DLBCL.

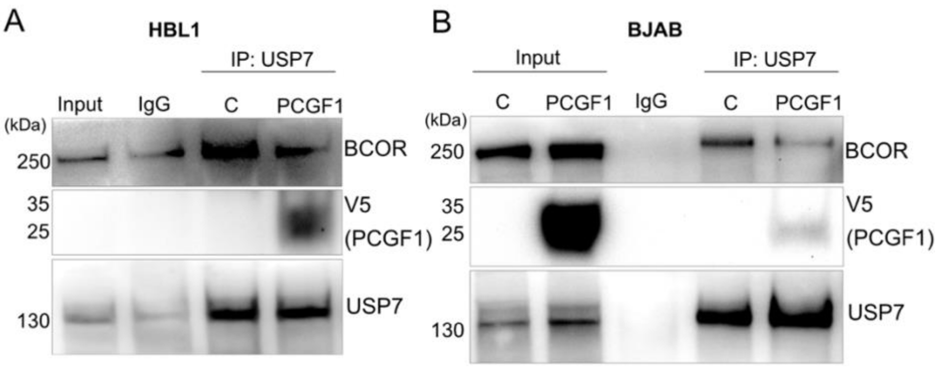

Figure 21: Endogenous USP7 interacts with PRC1.1 components in non-mutated DLBCL cell lines. (A) Co-immunoprecipitation in HBL1 cells showing the interaction between endogenous USP7 with BCOR and PCGF1 (V5 tag). Immunoprecipitation was performed using an anti-USP7 mouse antibody and IgG as a negative control. Western blot analysis was performed with antibodies against BCOR, V5 tag (PCGF1), and USP7. (B) Co-immunoprecipitation was performed in BJAB cells using the same approach.

To further understand which functional domains of USP7 are important for the interaction with PRC1.1 members, truncated and full-length HA-tagged versions of USP7 were generated (***Figure 22A*** and ***B***). The truncated versions of USP7 lacked either the N-terminal TRAF and UCH domains (ΔTRAF ΔUCH) or the C-terminal UBL domains (ΔUBL1-5) (***Figure 22C***). Next, I transduced PCGF1-overexpressing cells with the generated full-length and truncated USP7 constructs and performed immunoprecipitation against the HA tag to test the specificity of interaction between the PRC1.1 components. Only full-length USP7 showed interaction with V5-tagged PCGF1, as well as with BCOR, while the truncation variants failed to demonstrate such associations (***Figure 22C***). The findings that only full-length and no other construct was able to exhibit the interaction indicate that likely the 3D structure after folding is required for the interaction between USP7 and PRC1.1.

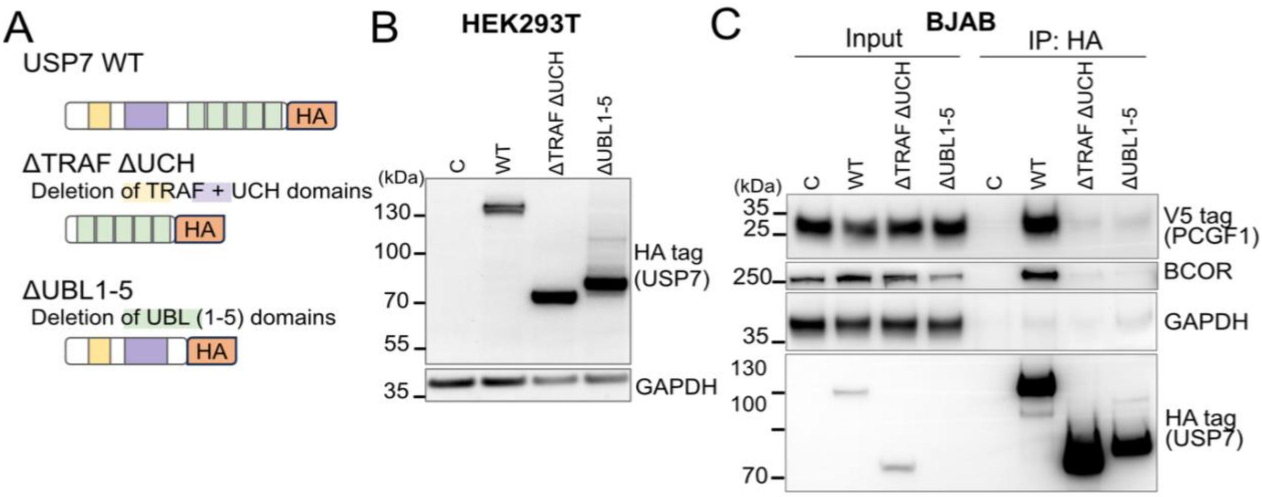

Figure 22: USP7 Interaction studies with PCGF1 and BCOR. (A) Cartoon representation of the USP7 constructs. (B) Validation of the overexpression of full length USP7 and truncated constructs in HEK293T. GAPDH was used as loading control. (C) HA tag (USP7) immunoprecipitation reveals that the interaction between USP7 and PRC1.1 members requires the protein in its full length.

3.10 USP7 mutations impair protein functionality

Having established the domain requirements for USP7 to interact with PRC1.1 subunits, it is important to acknowledge that patients harbor missense mutations in USP7 rather than complete deletions of functional regions. DLBCL patients were found to harbor 3 mutations around the N-terminal TRAF domain, 4 mutations in the catalytical UCH domain, 3 mutations in between UCH and UBL1-3 domains, 4 mutations in the UBL1-3 domains and 2 mutations in the UBL4-5 domains. To address the functional consequences of this widespread mutational pattern, four mutations were identified within distinct functional regions of *USP7*, namely the critical domains, UCH, UBL1-3, and UBL4-5 (***Figure 23***). Each of these mutations could potentially impact protein interactions and function. To this end, I investigated each of them for their potential effects.

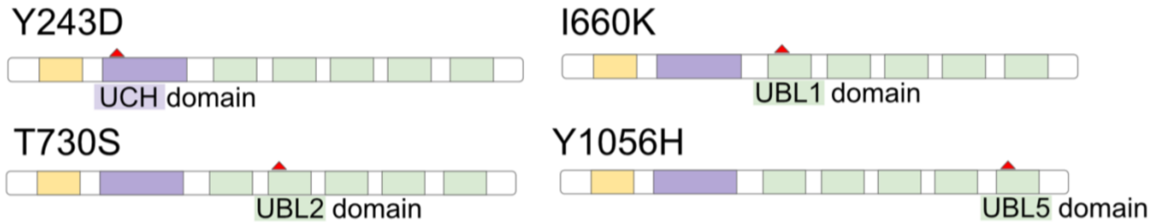

Figure 23: Mutations in USP7 from DLBCL patients assessed in the study. Cartoon representation of four missense mutations from DLBCL patients occurring in distinct domains from *USP7*. Y243D occurs in the UCH domain, I660K in the UBL1 domain, T730S in the UBL2 domain and Y1056H in the C-terminal UBL5 domain. Red triangle represents the approximate location of the respective missense mutation in the gene locus.

To understand the impact of the selected *USP7* mutations that can be found in primary DLBCL patients to the integrity of the PRC1.1 complex, I transduced parental K422 and HBL1 cells with four different HA-tagged USP7 mutations (Y243D, I660K, T730S, and Y1056H) cloned into a pLX312 vector containing GFP as selection marker. After transduction, cells were sorted based on GFP expression to ensure a pure population of cells, achieving a GFP positivity above 90% for subsequent analysis. Overexpression of the different USP7 variants was validated by checking RNA (***Figure 24A***) and protein levels (***Figure 24B***). Given the differences in overall protein levels of USP7 variants, I hypothesized that the different USP7 mutations affect the protein stability of USP7. For this reason, I treated HBL1 cells with cycloheximide, a tool compound blocking eukaryotic translational elongation, for 6 and 12 hours or left the cells untreated (***Figure 24C***). Notably, the mutated forms have, after 6-12h treatment with cycloheximide, a marked reduction in protein abundance, indicating that the USP7 mutations significantly reduce the half-life of the USP7 protein.

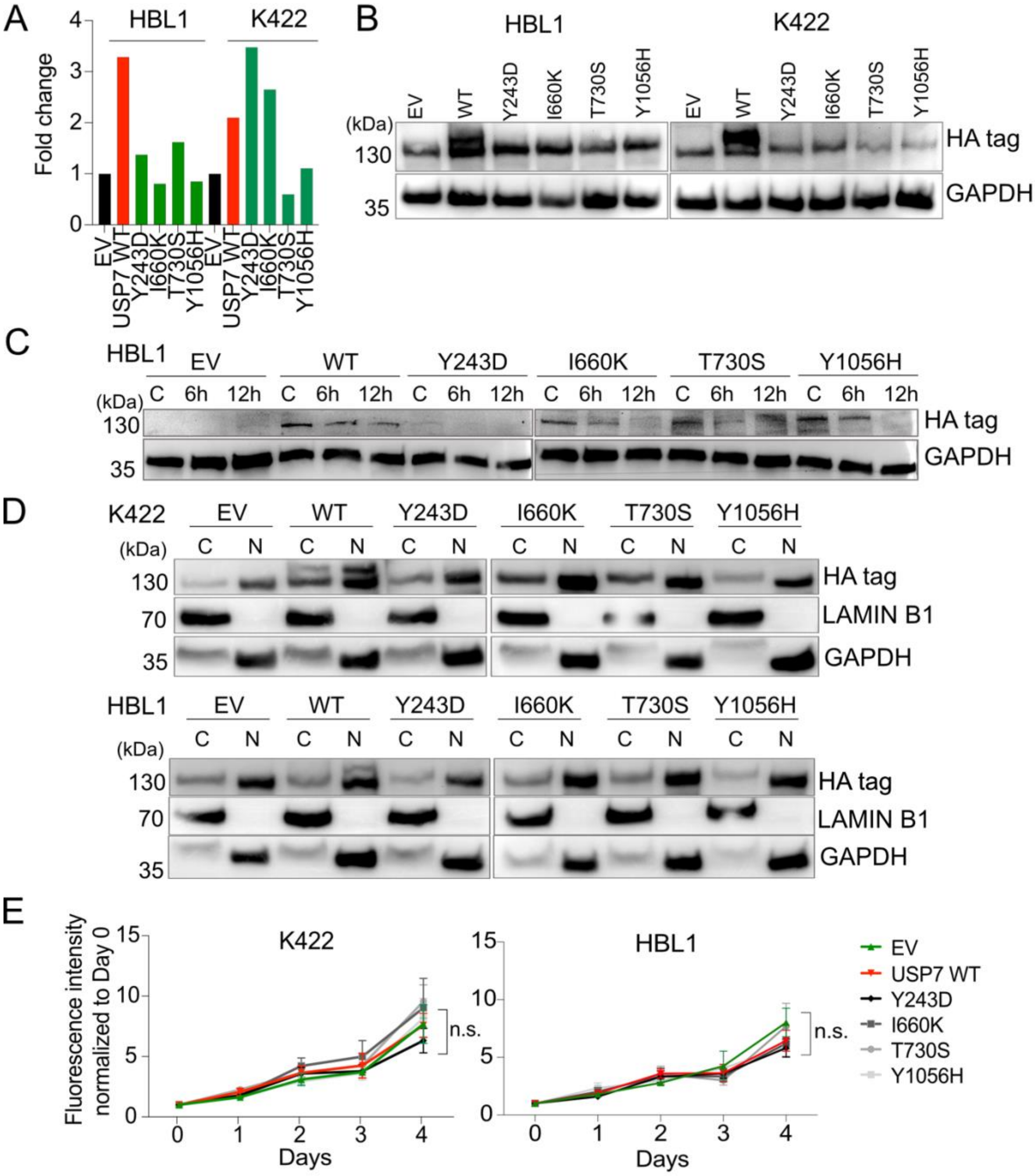

Figure 24: Expression and analysis of USP7 mutations. (A) RNA levels of USP7 in K422 and HBL1 cells overexpressing different versions of USP7 (WT and the Y243D, I660K, T730S and Y1056H mutations). GAPDH was used as the housekeeping gene for the normalization of CT values. (B) Confirmation of the expression of the HA-tagged USP7 mutations in K422 and HBL1. GADPH was used as loading control. (C) Immunoblot showing the degradation of WT USP7 and mutations upon treatment with cycloheximide (CHX) in HBL1 cells for 6 and 12 hours. "C" indicates control (not treated). (D) Proliferation of HBL1 and K422 cells overexpressing the different USP7 constructs in pLX312 (USP7 WT, Y243D, I660K, T730S and Y1056H) and an EV control. Proliferation was measured by fluorescence intensity measurement of the reduction of resazurin (alamarBlue) to resorufin and fluorescence intensity was normalized to Day 0. (E) Western blot depicting the subcellular localization of control (EV) and the USP7 versions (WT, Y243D, I660K, T730S, and Y1056H) in HBL1 and K422. Nuclear fraction is indicated as "N" and has LaminB1 as the nuclear and loading control, while the cytosolic one is indicated as "C", having GAPDH as the cytosolic and loading control for these cellular fractions.

To understand whether USP7 mutations also alter its cellular localization and potentially affects its function, subcellular fractionation experiments in K422 and HBL1 cells were performed. I observed that WT USP7 and its mutants accumulate in the nucleus and cytosol (***Figure 24D***), with a predominance for cytosolic localization, especially for the ubiquitin-bound WT USP7. In HBL1, no ubiquitin-bound WT USP7 is observed in the nucleus, while in K422 there seems to be a small portion accumulating in the nucleus. Mutations I660K and T730 have a higher amount of nuclear protein than other mutants. WT USP7 was strongly located in the cytosol, with an additional upper band signal, which indicates ubiquitinated USP7, when the protein has active deubiquitinating activity[85]. This further indicates that the USP7 mutations, lacking the upper band, possess some disruption in the deubiquitinase activity of the protein. In addition, I followed the proliferation of WT and USP7 mutations in K422 and HBL1 for 5 days, but I did not observe any significant difference when WT USP7 or mutations were expressed in these cells (***Figure 24E***), compared to the empty vector (EV) control.

Table 31: Pathogenicity predictions for USP7 missense mutations

Missense mutation	Score	Pathogenicity
I660K	0.9997	likely_pathogenic
Y243D	0.9996	likely_pathogenic
V570D	0.9994	likely_pathogenic
L865P	0.9993	likely_pathogenic
H156R	0.9991	likely_pathogenic
L910P	0.9975	likely_pathogenic
S363T	0.921	likely_pathogenic
Y1056H	0.6667	likely_pathogenic
D271E	0.4403	ambiguous
T730S	0.1374	likely_benign
A671G	0.1036	likely_benign
T63S	0.0826	likely_benign
T37S	0.0748	likely_benign
A530V	0.091	likely_benign

Additionally, the potential impact of USP7 missense mutations to be deleterious or benign (defined as pathogenicity) using AlphaMissense, an artificial intelligence (AI) model built on AlphaFold2, was assessed (***Table 32***). AlphaMissense gives a score (0-1) based on the likelihood of a missense mutation being benign (equals 0) or pathogenic (equals 1) (***Table 31***). Mutations such as I660K (0.9997), Y243D (0.9996), and Y1056H (0.6667) showed high scores associated with pathogenicity, while T730S had a low score (0.1374), predicted to likely be benign. Interestingly, these data are consistent with the cycloheximide experiments, in which T730S had the smallest effect.

3.11 USP7 mutations disturb interaction with PRC1.1 members

To evaluate the functional impact of USP7 mutations, I transduced K422 and HBL1 cells already expressing exogenous V5-PCGF1 with wild-type (WT) and mutated versions of USP7 and tracked cell proliferation over four days (***Figure 25A***). No significant differences in proliferation were observed between any of the mutants (Y243D, I660K, T730S and Y1056H) and the control (PCGF1/EV) in both K422 and HBL1 cells. This suggests that the introduced USP7 mutations did not notably affect the proliferation rate of these cell lines.

Next, I examined how these mutations affected interactions with PRC1.1 members by performing immunoprecipitation of WT and mutated USP7 variants. Followed by western blot analysis (***Figure 25B***). The results revealed a differential effect of these mutations on the interactions with BCOR and PCGF1. While T730S and Y1056H mutations in USP7 did not fully compromise USP7 binding interactions with BCOR and PCGF1, as evidenced by retained co-immunoprecipitation signals, Y243D and I660K mutations led to a complete loss of binding with BCOR and PCGF1, indicating a strong impact of these mutations on protein interactions.

These findings align well with the AlphaMissense data, where Y243D and I660K received high pathogenicity scores, suggesting a high likelihood of being deleterious. The complete loss of interaction between USP7 and PRC1.1 components for these mutations supports this prediction, indicating their disruptive nature. Conversely, T730S was predicted as benign, and its retained binding capacity aligns with the lower pathogenicity score that was observed. The Y1056H mutation, which showed an intermediate pathogenicity score, also retained binding to PRC1.1, consistent with a less severe but still impactful alteration. These experimental findings suggest that the disruption in interaction with PCGF1 and BCOR in Y243D and I660K mutants, could contribute to the protein’s dysfunctional role in DLBCL. On the other hand, the preserved or less severe interaction disruption to PRC1.1 members by T730S and Y1056H mutations, respectively, indicate a milder impact of these mutations on the USP7 functional role in DLBCL.

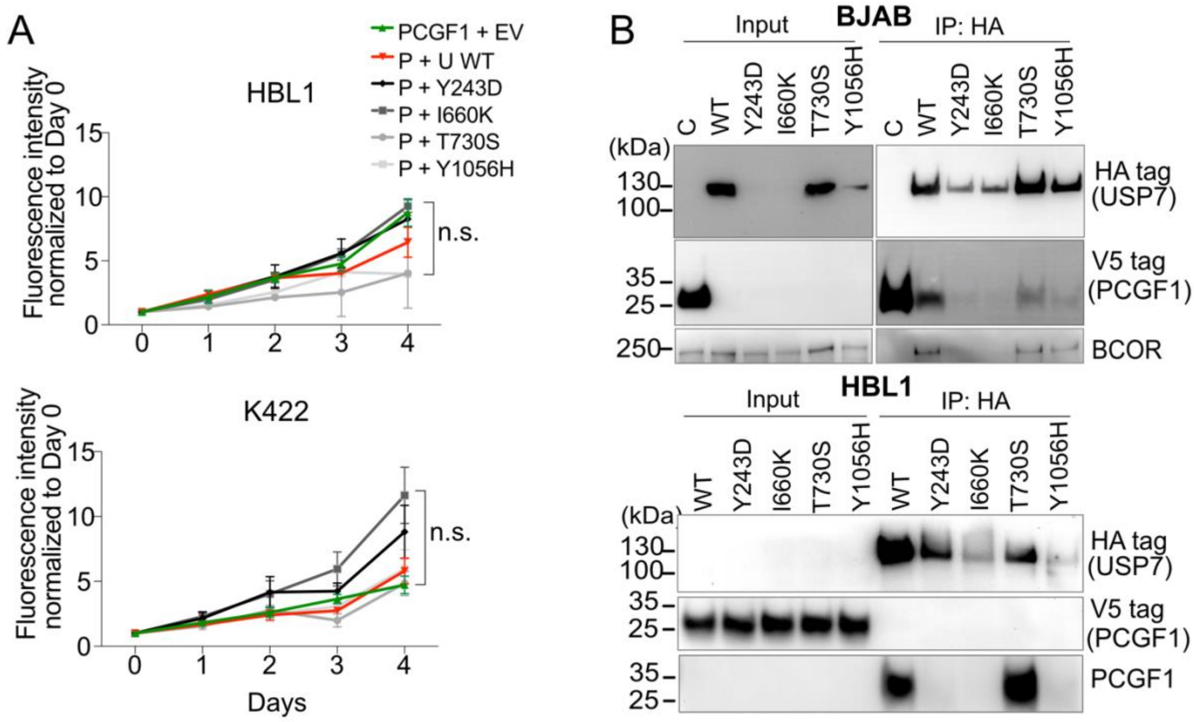

Figure 25: USP7 mutations impair binding to PRC1.1 subunits. (A) Proliferation of HBL1 and K422 cells transduced concomitantly with PCGF1 and WT and USP7 mutant versions. Proliferation was assessed with alamarBlue, as previously described. PCGF1 is control (PCGF1+EV). Fluorescence intensity was normalized to Day 0, as previously mentioned. (B) Immunoprecipitation of HA-tagged USP7 versions in BJAB (upper) and HBL1 (lower) cells reveals an impairment in the binding of the USP7 mutants Y243D and I660K to PCGF1 (V5 tag) and BCOR.

3.12 *In silico* prediction of protein structures and complex assembly in PRC1.1

To understand how mutations in *USP7* impact protein structure and the complex assembly in a three-dimensional (3D) way, *in silico* prediction tools were used in collaboration with Amal Hassan Ali from the Institute of Medical Physics and Biophysics in Charité, Berlin to investigate the interaction in detail. Crystal structures, when available, were obtained from the Protein Database (PDB). USP7, being one of the most well-characterized deubiquitinases, has several crystal structures covering distinct domains available of the human protein. However, there are no structures available covering it entirely. For BCOR, there are five crystal structures available on PDB, covering regions that are particularly important for protein-protein interactions, such as BCOR's PUFD domain that interacts with the RAWUL domain from PCGF1, as well as BCOR's BCL6-binding domain that interacts with BTB domain from BCL6. However, no full-length crystal structure of BCOR has been solved until now in humans, most likely due to the challenges posed by its long and dynamic sequence structure composed of 1755 amino acids. A way to overcome the lack of crystal structures available is to generate protein structure prediction using artificial intelligence. The predictions of BCOR and USP7 structures were performed using Robetta and AlphaFold. The full-length USP7 protein structure was obtained from the AlphaFold protein database (AF-Q93009-F1-v4) and its structural similarity to USP7 was assessed using separate crystal USP7 structures from PDB which, in combination, cover the entire protein. Three different crystal structures were aligned to the AlphaFold USP7 predicted structure (AF USP7): 5J7T, which covers the catalytic domain and parts of the UBL domain of the protein; the N-terminal TRAF domain from USP7 deposited as 2F1W, as well as the C-terminal crystal structure 7VIJ (***Figure 26***). These sequences were aligned in overlapping areas using PyMOL and the root mean square deviation (RMSD) was calculated for each structure alignment (***Table 32***). These RMSD values indicate a good degree of structural conservation between the predicted and experimental structures, particularly in the aligned regions of the N-terminal TRAF.

Table 32: Alignment of the USP7 crystal structures to the predicted alpha fold structures

PDB ID	Domain	Number of aligned atoms	Root Mean Square Deviation (RMSD)
2F1W	N-terminal TRAF	116	0.368 Å
7VIJ	C-terminal UBL1-5	508	8.650 Å
5J7T	UCH + UBL1-5	659	6.424 Å

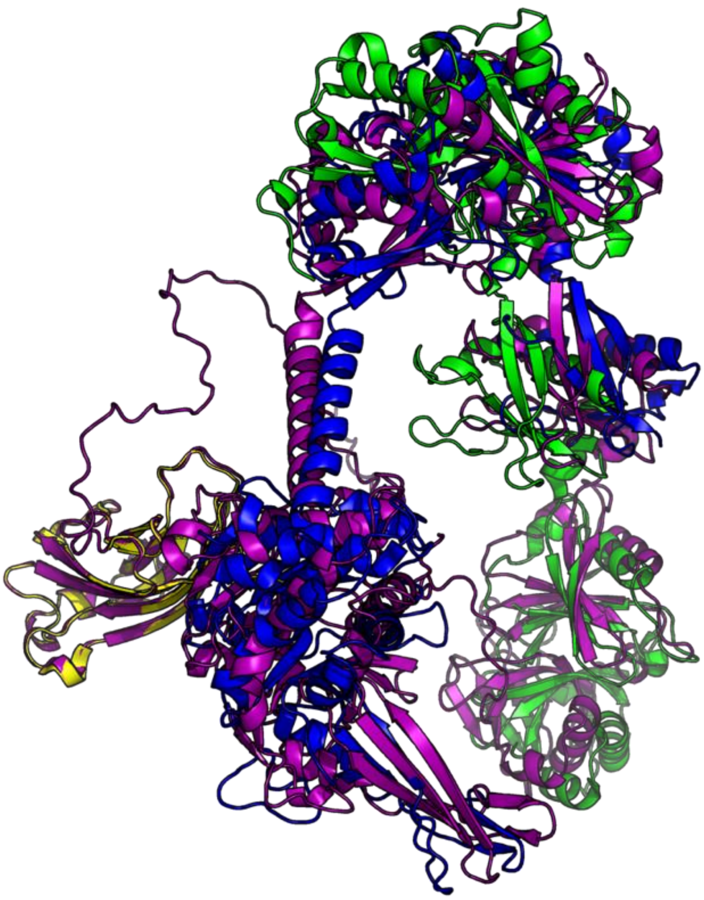

Figure 26: Superimposition of the AlphaFold-predicted model of USP7 and USP7 crystal structures from PDB. The AlphaFold-predicted model of USP7 (purple) was superimposed to crystallized structures from USP7: TRAF domain (PDB ID: 2F1W, yellow), UCH and UBL1-5 domains (PDB ID: 5J7T, blue) and UBL1-5 domains (PDB ID: 7VIJ, green). AlphaFold prediction was done by Amal Ali. The superimposition figure was generated using PyMOL by D. Joppi.

The predicted USP7 sequence was colored purple. In addition to that, the mutated amino acid residues (missense mutations) in USP7 from DLBCL patients were highlighted in green in this structure for better visualization of the affected domains in 3D perspective (***Figure 27*** and ***Figure 28***).

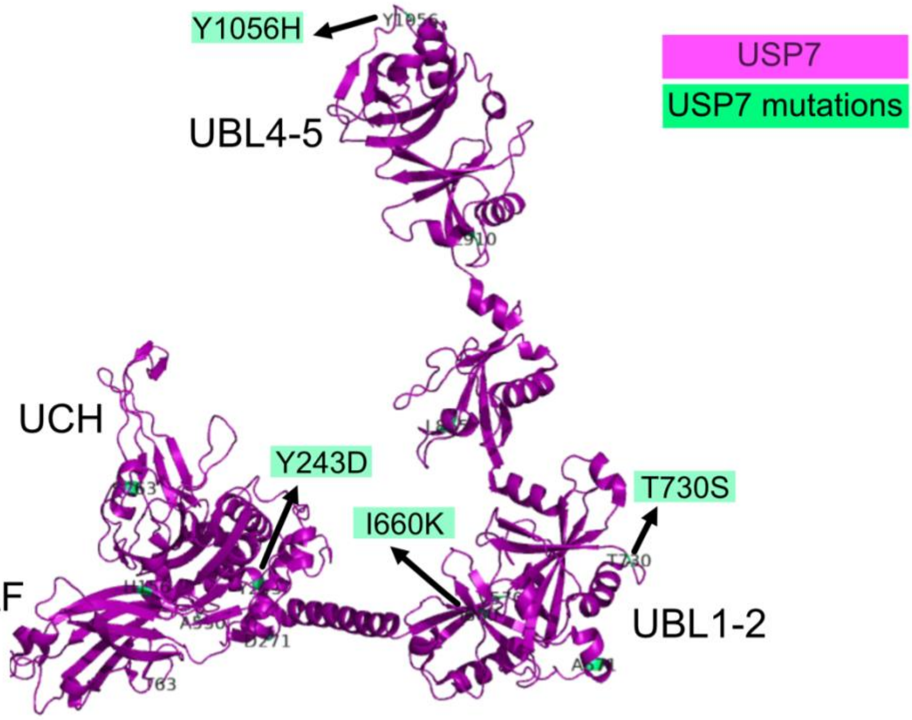

Figure 27: Mapping of mutation sites in USP7. Structural representation of USP7 structure, showing the domains (TRAF, UCH, UBL1-5). Mutations in USP7 (green) are mapped onto the structure to show their spatial distribution. The color scheme represents the combined USP7 sequence (magenta) and highlights the mutation sites in green. AlphaFold prediction was done by Amal Ali. Figure was generated using PyMOL by D. Joppi.

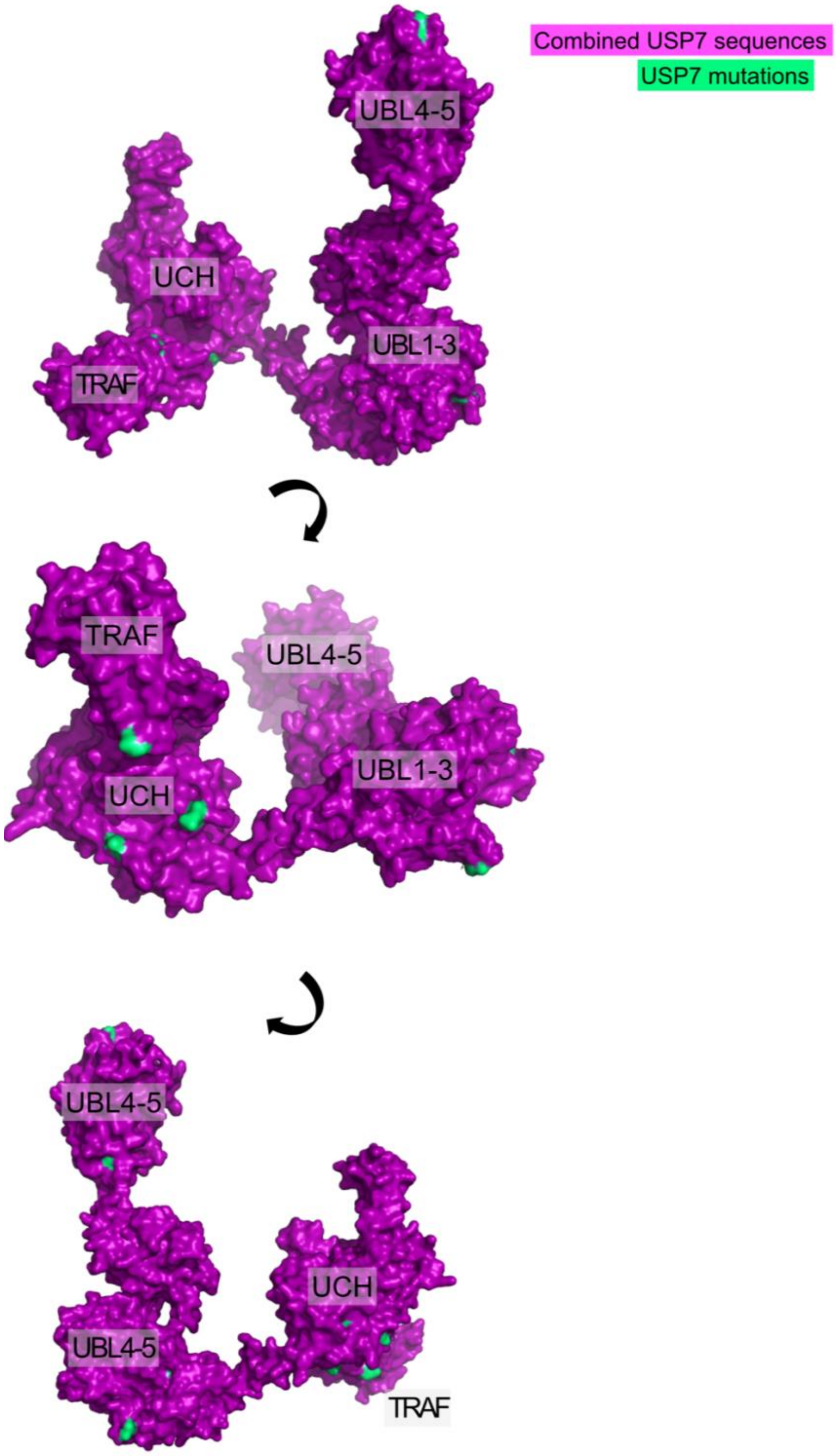

Figure 28: Surface representation of USP7 structure and mapped mutations. Three views (90° rotations) of the USP7 structure is shown, represented in surface mode, highlighting distinct domains and mutation sites. The structure is colored according to the combined USP7 sequences (magenta) with known mutation sites mapped in green. The domains include TRAF (tumor necrosis factor receptor-associated factor), UCH (ubiquitin carboxyl-terminal hydrolase catalytic site), and UBL (ubiquitin-like) domains UBL1-3 and UBL4-5. Green highlights indicate mutations in USP7, which are distributed across different domains, particularly near the catalytic UCH domain and within the UBL domains. The three views represent rotational orientations to better visualize the spatial relationships and mutation positions across the structure. AlphaFold prediction was done by Amal Ali. Figure was generated using PyMOL by D. Joppi.

The interaction between BCOR and USP7 was predicted using AlphaFold and this prediction (in yellow) was aligned to the combined USP7 sequence (purple) (***Figure 29A***). It is possible to observe that the predicted binding regions of BCOR and USP7 occur in the N-terminal of USP7 (regions containing UCH and TRAF domains). When zooming in on the N-terminal USP7 region, it is evident that mutations (in green) affecting this region, such as Y243D, will likely affect the binding between BCOR and USP7 ***(Figure 29B)*.**

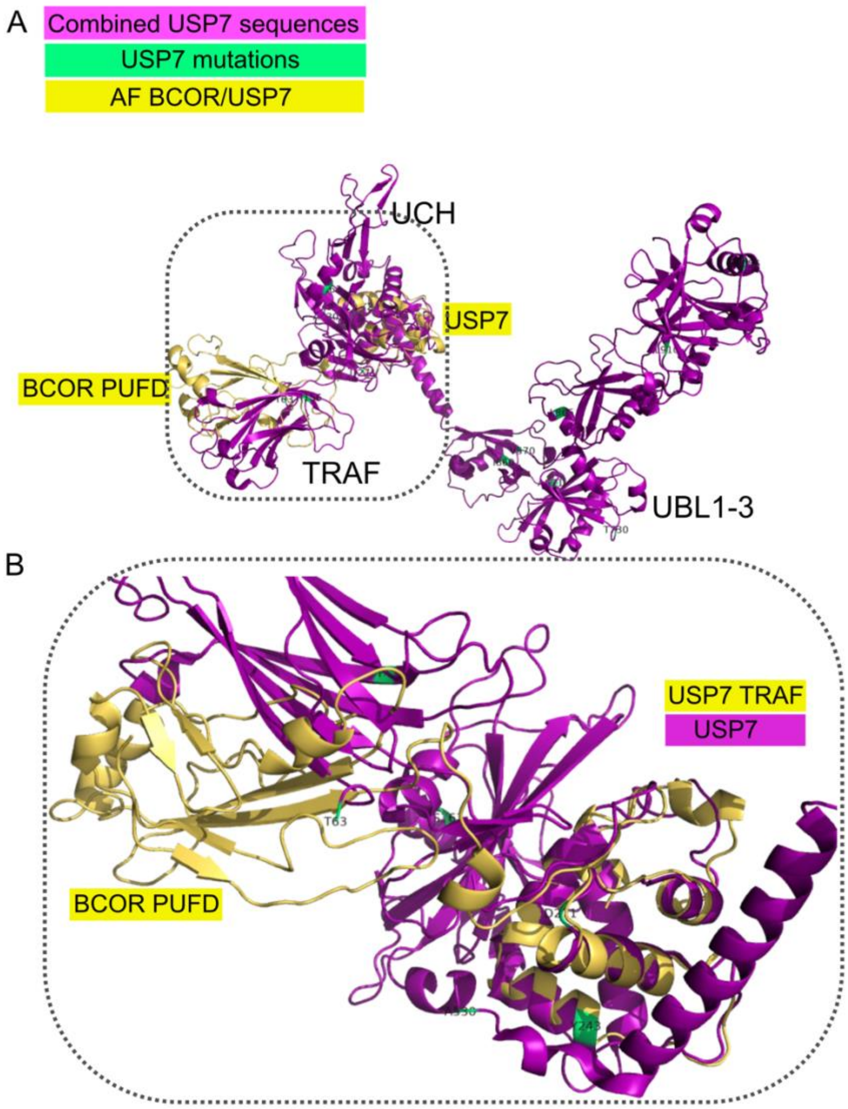

Figure 29: Structural model of BCOR interaction with USP7 and mapped USP7 mutations. (A) Overview of the combined USP7 structure (magenta) and the AlphaFold-predicted BCOR PUFD domain (yellow), highlighting their interaction. The main USP7 domains are labeled, including the TRAF (tumor necrosis factor receptor-associated factor), UCH (ubiquitin carboxyl-terminal hydrolase), and UBL1-3 domains. USP7 mutations (green) are distributed across the structure, with key residues near the BCOR interaction site. (B) Zoomed-in view of the interaction between the BCOR PUFD domain and the USP7 TRAF domain, showing structural details of the interface. Mutation sites in USP7 are labeled in green. The BCOR PUFD domain is positioned near the UCH domain of USP7, suggesting a potential regulatory interface. Colors are as follows: USP7 structure (magenta), BCOR PUFD (yellow), mutations (green). AlphaFold interaction prediction was done by Amal Ali. Figure was generated using PyMOL by D. Joppi.

For the full-length structural prediction of BCOR, the amino acid sequence was divided into three parts, to allow a more precise prediction on Robetta. Thus, three segments were obtained (***Figure 30A***, ***B***, and ***C***). One of the few available crystal structures of BCOR (PDB ID: 2N1L), which includes the PUFD domain, was aligned to the AlphaFold-predicted structure of BCOR's C-terminal segment (***Figure 30C***). The alignment produced an RMSD of 3.134 Å across 82 atoms, suggesting that the overall fold of the two structures is reasonably well-aligned. However, minor local deviations are present, reflecting both structural flexibility and the limitations of predictive modeling. These three sequences were combined and the full-length BCOR structure was generated (***Figure 30D***), which allows us to have a better understanding of the structural distribution and dynamics of this protein.

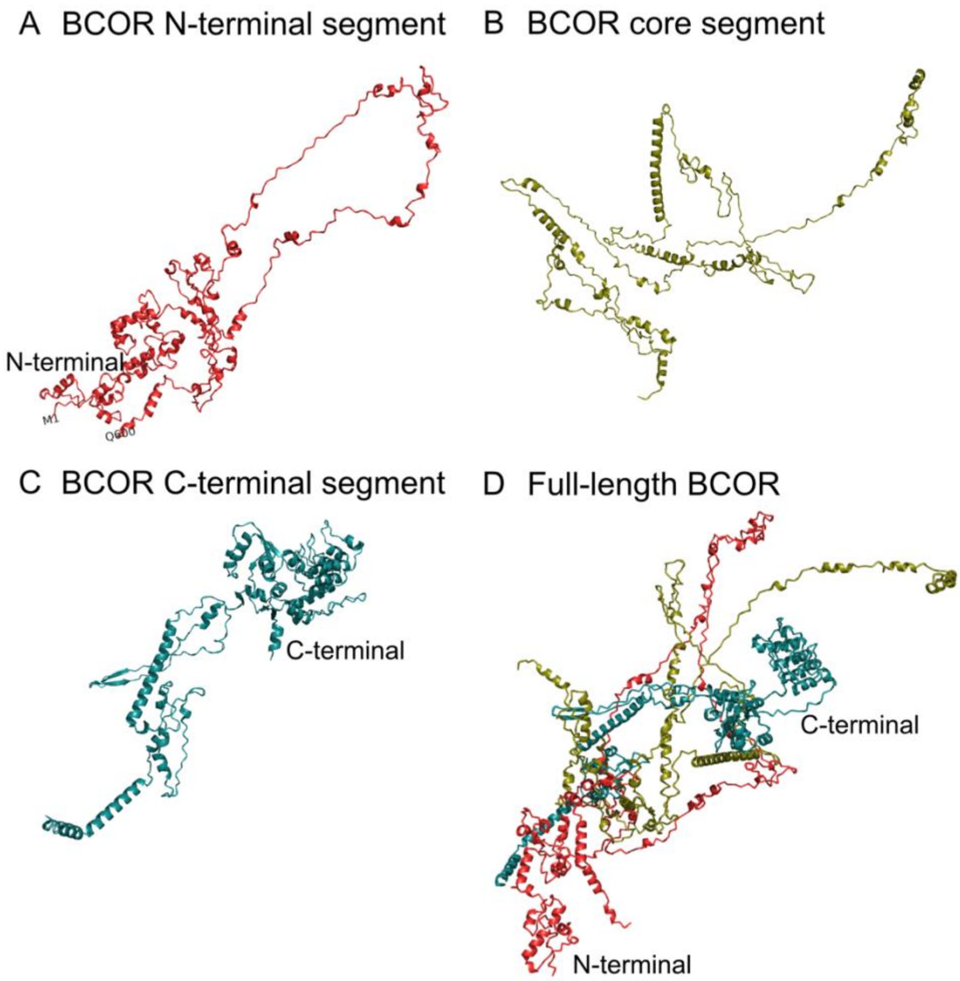

Figure 30: Predicted structural segments and full-length model of BCOR. (A) The N-terminal segment of BCOR, shown in red, includes the N-terminal domain and disordered regions from amino acids 1 to 600. (B) The core segment of BCOR, highlighted in yellow, consists of helices and loop regions, representing the central portion of the protein from amino acids 601 to 1185. (C) The C-terminal segment of BCOR, shown in teal, forms a distinct folded structure and covers amino acids 1186 to 1755. (D) The full-length BCOR AlphaFold predicted structure, combining all three segments (N-terminal in red, core in yellow, and C-terminal in teal), illustrates the overall arrangement of the N- and C-terminal domains with the core segment. BCOR protein structure predictions was done by Amal Ali. Figures were generated using PyMOL by D. Joppi.

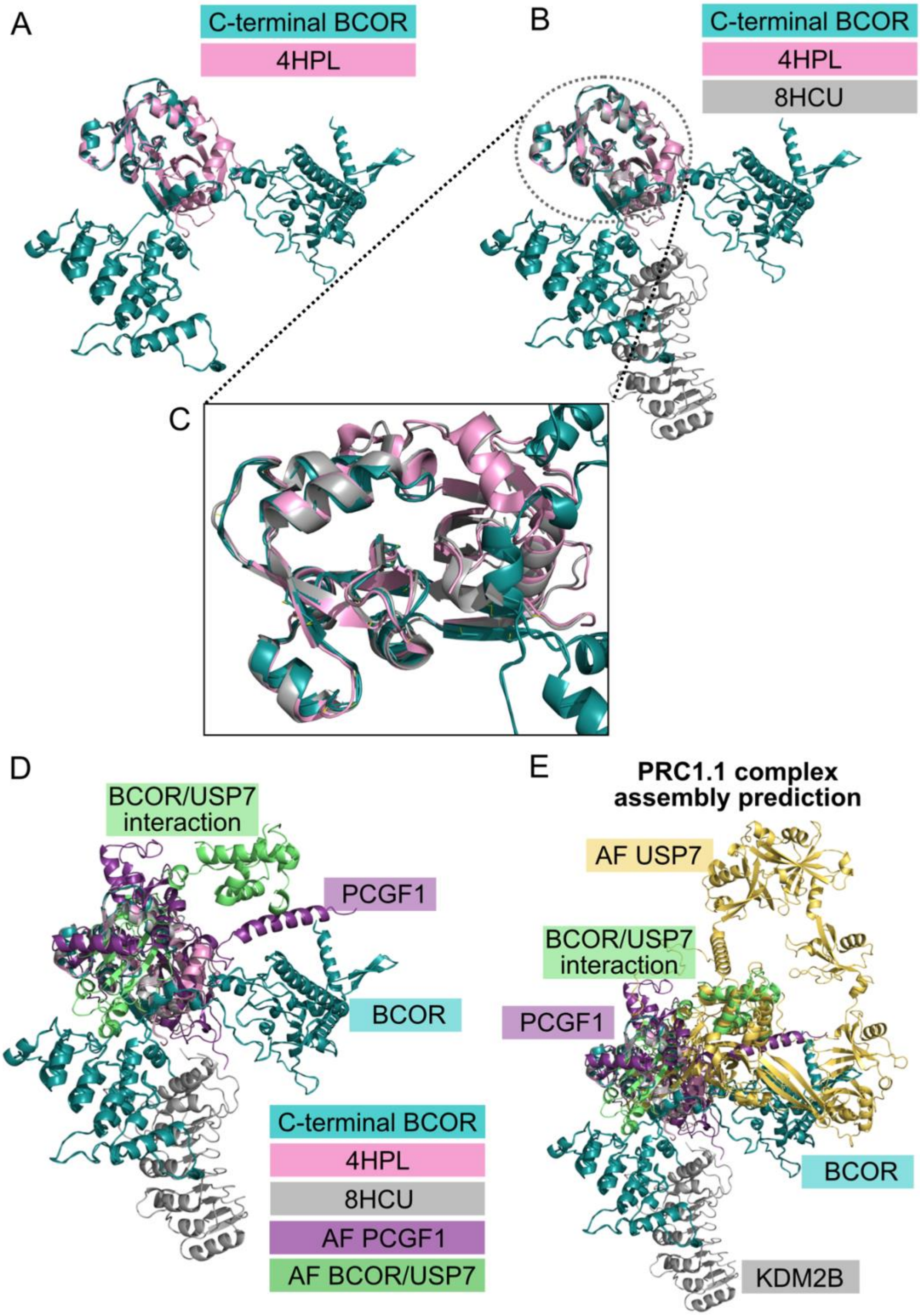

Figure 31: Three-dimensional modeling of the PRC1.1 complex. (A) Overlay of the predicted of the C-terminal BCOR domain (teal) with the crystallized structure 4HPL (pink) to validate prediction accuracy. (B) C-terminal BCOR overlaid with 4HPL (pink) and 8HCU (gray) in proximity and magnified view in (C). (D) Overlaid structures from (C) with predicted PCGF1 structure (purple) and BCOR:USP7 interaction site (green). (D) Assembly predition of the complete PRC1.1 complex with structures from (D) added with AlphaFold USP7 structure (yellow). *In silico* predictions were done by Amal Ali. Figures were generated using PyMOL by D. Joppi.

Additionally, the alignment of the PCGF-BCOR binding crystal structure (PDB ID: 4HPL) to the predicted structure of BCOR's C-terminal was successfully done (***Figure 31A***) with a RMSD of 0.938 Å across 98 atoms. The alignment of the crystal structure of the BCOR/PCGF1/KDM2B complex (PDB ID: 8HCU) was also performed, showing a well-aligned fit to the C-terminal BCOR region (***Figure 31B*** and ***C***) with a RMSD of 0.732 Å across 102 atoms.

The comprehensive modeling of the PRC1.1 suggests a reliable conformation, where the C-terminal region of BCOR is superimposed with 4HPL and 8HCU, showing consistent structural similarity (***Figure 31C***), as well the overlay with AlphaFold-predicted PCGF1 and USP7:BCOR structures (***Figure 31D***). The overlay and alignment of the AlphaFold-predicted USP7 structure indicates that all the members assemble in reliable conformation, where the C-terminal of BCOR is superimposed with 4HPL and 8HCU, and USP7 and PCGF1 are performing contacts (***Figure 31E***). These findings have significant implications. Firstly, they confirm that our predicted structures and the assembly of PRC1.1 subunits are consistent with experimental crystallographic data. Secondly, this provides for the first time, a 3D model of the PRC1.1 complex, which includes BCOR, PCGF1, USP7 and KDM2B.

4 Discussion

4.1 Identification of PRC1.1 mutations in DLBCL

In this work, I comprehensively identified somatic mutations in all Polycomb repressive complexes using 878 *de novo* DLBCL patients from two independent cohorts. This analysis identified recurrent somatic mutations in the non-canonical PRC1.1 complex subunits, *USP7*, *BCOR*, *BCORL1,* and *KDM2B,* compared to other PRCs: 37 mutations in Chapuy *et al*. and 56 in Schmitz *et al.* (***Table 27***). Notably, these mutations did not occur in a specific DLBCL molecular subtype, nor did they co-occur with each other or with gain-of-function PRC2 mutations (*EZH2* Y641X), suggesting a distinct mode of PRC1 deactivation in DLBCL. Mutations in *USP7*, *BCOR*, *BCORL1,* and *KDM2B* were predominantly missense and were distributed across the open reading frame, particularly in *USP7*, which had mutations occurring in all domains, indicative of a role as a putative tumor suppressor. An enrichment of *USP7* mutations was found specifically in transcriptionally defined GCB-type DLBCLs and in the corresponding genetic C3 and C4 DLBCLs, indicating the biological relevance of the mutations to these subtypes. To investigate the implications of these patient data and pick relevant DLBCL model systems, I analyzed and chose from 28 DLBCL cell line models a representative panel for my studies. These cell lines exhibited similar patterns of PRC1.1 subunit mutations as identified in primary patients, underscoring that they can be used as informative models.

4.1.1 *USP7* mutations

Among the PRC1.1 members, *USP7* was one of the most mutated subunits with 19 mutations detected in DLBCL patients (***Figure 10A***). Based on the observed mutational pattern, I hypothesized that *USP7* mutations likely led to loss-of-function since the dispersed mutations occurred in all domains of USP7. The affected domains were essential for the proper enzymatic function of the protein or interaction with other proteins. In the catalytical UCH domain of USP7 (from residue 208 to 537) (***Figure 11***), 5 mutations were observed in patients, with 4 out of 5 being missense: Y243D, D271E, S363T, and A530V. In this domain, ubiquitin molecules bind to residue C223, and this interaction allows the UCH domain to cleave the isopeptide bond between ubiquitin bound to lysine residues from the target protein of USP7[102]. Besides, two other residues, H464 and D481, are part of the catalytic triad, pointing out their relevance as well in terms of contact. Thus, a disturbance in the protein activity with these three different mutations affecting the UCH domain (Y243D, D271E and S363T) occurring in patients would be expected. From the AlphaMissense analysis (***Table 31***), however, only Y243D and S363T were predicted to likely be pathogenic. USP7 has also been described to have a long alpha-helix from residues 538 to 562 that connects the UCH and UBL domains and allows conformational changes that regulate the activity of the protein[109]. The UBL domains have also been shown to be critical for the regulation and full enzymatic function of the protein[102]. These UBL domains are organized in a 2-1-2 pattern (UBL1-2, UBL3 and UBL4-5). The UBL4-5 domains, in particular, are crucial for the correct folding of the C-terminus into the catalytical pocket, thus guaranteeing the catalytical activity of the protein[102]. In DLBCL patients, I have identified mutations occurring in all the domains: V570D in the connecting alpha-helix; Q617*, I660K, A671G in UBL1-2; T730S and L865P in UBL3; and L910P and Y1056H in UBL4-5. Of these missense mutations, four out of six are predicted to likely be deleterious (***Table 31***). Similarly, the Q617* nonsense mutation would likely lead to a non-functional protein due to its disruption of UBL domains. These mutations thus suggest impaired protein function and disrupted interactions with target proteins, which is mediated by UBL domains and the N-terminal TRAF domain. The TRAF domain is unique for USP7 among the deubiquitinases from the ubiquitin-specific protease (USP) family, and in DLBCL patients, two mutations were observed in the N-terminal, but both being predicted to be benign (***Table 31***).

4.1.2 *BCOR* mutations

Mutations in *BCOR* (n = 22) were also recurrent in DLBCL patients and showed a tendency to be enriched in C5 DLBCL patients (p-value = 0.067), which is an ABC-type-enriched DLBCL cluster. Like *USP7*, most of the mutations in *BCOR* were missense. However, a total of six truncating mutations were seen in *BCOR*, predominantly affecting the C-terminal region of the protein. It is important to consider that *BCOR* is located in the X chromosome; thus, truncating or loss-of-function mutations in male patients likely lead to a complete loss of function, while in female patients, the extra allele in females could potentially compensate for the loss, resulting in different outcomes, besides both presenting deleterious mutations. In addition, a similar pattern of truncating mutations in *BCOR* and *BCORL1* has been seen in leukemia. These mutations, which affect the C-terminal of both BCOR and BCORL1, are highly frequent and disruptive for the PRC1.1, preventing these subunits from binding to the RING1A/B-PCGF1 core[92]. In this context, however, *BCOR* and *BCORL1* are the only mutated PRC1.1 subunits, and they were shown to form independent PRC1.1 complexes (BCOR-PCGF1 and BCORL1-PCGF1, respectively), but with shared target loci[92], also indicating the possibility of PRC1.1 to form subcomplexes with overlapping functions. In DLBCL, the remaining and at the same time majority of mutations in patients were concentrated around the BCL6-binding domain (BBD) as well as in the ANK/PUFD domain regions. While mutations in BBD can disturb the interaction of the protein with the BCL6 dimer and thus the BCOR co-repression, mutations in the ANK and PUFD regions can impair BCOR interaction with KDM2B and PCGF1, respectively. Despite the functional role of ANK domain not being fully known, a region between ANK and PUFD named "linker" has been described as essential for the interaction between BCOR and KMD2B[86]. In fact, the minimal region necessary in BCOR for assembly of PRC1.1 with PCGF1, KMD2B, and SKP1 would cover residues 1564 to 1748 in the C-terminus, comprising the "linker" and PUFD domain. One particular mutation of a DLBCL patient was identified in the PUFD domain: L1664P, which occurs very close to the L1705 residue in the PUFD domain that has been shown to interact with KDM2B, apart from PCGF1[86]. However, this interaction with KDM2B can only occur upon the acquisition of a more PUFD hydrophobic state that allows hydrogen bond contacts, which takes place once PCGF1 binds and induces conformational changes of this BCOR region[87]. Thus, this mutation would likely lead to a disruption in the interaction between BCOR, PCGF1, and consequently KDM2B, having a pathogenicity score of 0.998 in AlphaMissense. Additionally, four cases harbored the hotspot N1459S mutation close to the ANK domain, which has been found in various types of cancer, including endometrial and ovarian, and predicted to be oncogenic[110]. However, its functional consequences are still unknown, but it may contribute to the interaction with other PRC1.1 members.

4.1.3 *BCORL1* mutations

Mutations seen in *BCORL1* were less numerous than in *BCOR* and *USP7*; however, they also have some accumulation in the N-terminal, a pattern similar to *BCOR*, containing ANK and PUFD domains as well. Similarly, to BCOR, BCORL1 also relies on its PUFD domain for interaction with PCGF1, and this interaction is required for subsequent binding to KDM2B[110]. Six mutations are seen in these domains, two being truncating mutations, which would likely compromise the proper functioning of the protein as part of PRC1.1. The missense mutations E1405K, P1590A, and S1543L are predicted in AlphaMissense to be benign, thus indicating that truncating mutations in BCORL1 would impact this protein's function and interaction with PRC1.1 members.

4.2 Role of PRC1.1 in B-cell development and hypothesized tumor suppression in DLBCL

The PRC1.1 complex has been reported to be atypically tethered to the chromatin in the normal development of B cells through CBX8 by PRC2 (EZH2) and BCL6, acting together with these factors to repress gene expression and maintain proper epigenetic control[59]. PRC1.1 subunits, in fact, get upregulated in germinal center B compared to naïve B cells (***Figure 13A***), indicating their role in maintaining germinal center functions. However, in DLBCL, mutations affecting PRC1.1 components likely compromise this regulatory function aided by PRC1.1, leading to gene expression dysregulation. This is particularly evident when the PRC1.1 mutations I observed do not co-occur with EZH2 as one would expect (***Figure 10C***). The hyperactivity of EZH2 in *EZH2*-mutated tumors likely needs some aid from PRC1.1, which is not required in other DLBCL subtypes where EZH2 does not play a major role as an epigenetic driver.

4.3 USP7 perturbation

To assess USP7's role in cell survival, I generated USP7 knockouts in the GCB-type DLBCL cell lines, LY1 ($EZH2^{mut}$) and LY18 (EZH^{wt}). The knockout did not significantly affect proliferation, and the orthogonal validation of these findings with a pharmacological inhibition of USP7 produced comparable results: DLBCL cell lines were not dependent concerning survival in our *in vitro* model on the structural function of USP7 or the enzymatic activity as a deubiquitinating enzyme. This finding differs from what has been seen in leukemia cells, where USP7 acts as an oncogene and its inhibition leads to cell death[111]. This supports the idea that in the context of DLBCL, USP7 may act as a tumor suppressor. In addition, I have observed that a subset of USP7-inhibited cell lines exhibited a loss of the H2AK119Ub mark, indicating an important PRC1-related effect upon USP7 disruption, which points to a connection between USP7 and PRC1 in this DLBCL.

4.4 Investigating the assembly of PRC1.1 in DLBCL

The next step was to investigate the assembly of the wildtype and mutated PRC1.1 multiprotein complex and used for these studies our DLBCL model cell lines. Immunoprecipitation of USP7 in parental lines revealed no detectable interaction with PCGF1, which could be explained by its low expression levels seen in DLBCL cell lines. Thus, hypothesizing that the low levels of PCGF1 consisted another mechanism to perturb PRC1.1 in DLBCL, I overexpressed PCGF1 alone in a panel of GCB (CTB1, LY1, LY4, BJAB, DHL5, K422 and HT) and ABC-type (TMD8, LY10, U2932 and HBL1) DLBCL cell lines containing PRC1.1 mutations (CTB1, LY4, HT, LY10) or not containing (the remaining cell lines). I observed that the PRC1-specific histone mark H2AK119Ub increased upon PCGF1 overexpression, particularly in cells non-mutated for PRC1.1 members, such as BJAB and HBL1 (GCB and ABC-type, respectively). Moreover, PCGF1 was localized in the nucleus, together with USP7, indicating that PRC1.1 could assemble in the same compartment under these conditions. Immunoprecipitations of endogenous USP7 in cells with *PCGF1* overexpression confirmed interactions with BCOR and PCGF1, showing successful PRC1.1 assembly. Furthermore, I treated PCGF1-overexpressing cell lines with a USP7 inhibitor and observed a decrease in the PCGF1-induced H2AK119Ub levels in ABC-type DLBCL cell line HBL1 or a complete loss in the GCB-type K422 and LY1 cell lines, suggesting that USP7 activity is required to sustain this histone mark. Next, to understand the domains or regions of USP7 involved in this interaction, I transduced PCGF1-overexpressing cells with different USP7 constructs (WT, ΔTRAF ΔUCH, and ΔUBL1-5). By pulling down HA-tagged USP7 versions, I saw that the interaction of USP7 with BCOR and V5-tagged PCGF1 only occurred with full-length WT USP7 was expressed, suggesting that the protein needs to have all domains and proper folding for the binding with PRC1.1 subunits to occur[112].

4.5 Expression patterns of PCGFs in DLBCL

One of the findings in this work was the low expression of *PCGF1* in DLBCL cell lines. *PCGF3* and *PCGF5* were observed to be expressed at high levels when compared to core canonical PRC1 subunits *PCGF2* and *PCGF4* (***Figure 16A***). In physiological GC B-cells, *PCGF1* is upregulated together with PRC1.1 subunits[59], yet its reduced expression in DLBCL further supports the notion that the PRC1.1 complex is impaired. Conversely, the high expression of *PCGF3* and *PCGF5* suggests that these subunits may compensate for the loss of *PCGF1*. However, this compensation may not be sufficient to fully replace PRC1.1's repressive activity, particularly in the tumor microenvironment of DLBCL. This differential expression of PCGF subunits points to a degree of functional redundancy within the PRC1 family but also raises the question of whether the specific loss of *PCGF1* contributes to the pathogenesis of DLBCL by altering the balance of chromatin regulation in favor of oncogenic gene expression.

On the other hand, PCGFs - and thus, PRC complexes - have been shown to have specific and unique target genes. PCGF1 has been described to possess 40% unique targets and shared targets with PRC1.2 (20%) and PRC1.2 or PRC1.6 (33%), while PCGF3 has been shown to have 47% unique targets and 14% common targets with PRC1.6[113]. Despite the low levels of *PCGF1*, it is still plausible that the stoichiometry of PRC1.1 may allow sufficient complex formation under physiological conditions. However, the mutations in PRC1.1 subunits may exacerbate PRC1.1 dysfunction. By providing enough PCGF1, through its overexpression, I ensured this subunit is available for binding to other PRC1.1 members. In fact, the assembly of PRC1.1 is hierarchical: PCGF1 first binds to BCOR, thus changing the disorganized structure of the latter, which is then able to bind to KD2MB, the DNA-binding subunits of PRC1.1[86]. This is another indication that, without appropriate levels of PCGF1, PRC1.1 in the subunit composition that we know it, is unable to assemble.

4.6 Unexplored roles of other PRC1 subcomplexes

Other Polycomb repressive complexes have not been addressed in this study; however, their importance is appreciated.

4.6.1 PRC1.3/1.5

For PRC1.3 and PRC1.5, a total of 68 mutations were found in primary patients. In addition, the core components, PCGF3 and PCGF5, respectively, are expressed at high levels in DLBCL cell lines, while PCGF2, PCGF4, and PCGF6 are expressed at low levels (***A***). BCOR has been described to have a PUFD domain that can bind both PCGF1 and PCGF3[114]. The fact that PCGF3 levels are high in DLBCL could indicate that an assembly of a complex containing BCOR-PCGF3 is not unlikely, suggesting that BCOR may still function independently of PRC1.1 in DLBCL, when not mutated. In this study, I did not find an interaction between USP7 and PCGF3 in co-IPs (data not shown); however, this does not exclude the possibility of BCOR still binding to PCGF3 in other cellular contexts, but this complex would likely then be USP7-independent.

4.6.2 PRC1.6

There is growing evidence that other PRC1 subcomplexes may also play a critical role in tumor suppression in DLBCL. One of the most notable components of PRC1.6 is MGA, a transcriptional repressor that has been described as a tumor suppressor and a driver gene for DLBCL, in which *MGA* mutations were seen to be enriched in the ABC-type[115]. Notably, my study is in line with the high frequency of *MGA* mutations in primary DLBCLs (***Figure 9C***). However, it is still uncertain whether this tumor suppressor function of MGA is PRC1.6-dependent. MGA has been observed to play a role in repressing MYC-driven oncogenesis[116]. In general, the tumor-suppressive role of MGA remains controversial, as demonstrated by the absence of significant lymphomagenesis in Eμ-Myc mouse models lacking MGA. The Eμ-Myc model, which is typically used to study Myc-driven B-cell lymphomas, did not show a clear tumor suppressor function for MGA. It is important to note that Eμ-Myc mice may not faithfully replicate the genetic and phenotypic traits of DLBCL, suggesting that the mechanisms underlying MGA's role in DLBCL may differ from other B-cell lymphoma models. The discrepancy between the *in vivo* findings in Eμ-Myc mice and the observed mutation frequency in ABC-type DLBCL patients underscores the complexity of MGA's function in DLBCL. Of note, *MGA* was also recently found to be frequently mutated in another form of aggressive B-cell lymphoma, Richter transformation of chronic lymphocytic leukemia to DLBCL[117]. It raises questions about whether MGA serves as a context-dependent tumor suppressor, with its role potentially varying across different DLBCL subtypes and genetic backgrounds. Interestingly, this context-dependent tumor suppressor role parallels what has been observed in PRC1.1. The subunits in PRC1.1 can exhibit divergent functions depending on the cellular context and genetic background, which are not universal but rather influenced by specific mutations, subtypes, or broader epigenetic landscapes. Thus, understanding how mutations in PRC1.1 and PRC1.6 components influence the epigenetic environment of DLBCL could inform the development of novel therapeutic strategies aimed at restoring PRC1-mediated gene repression.

4.7 Impact of PRC1.1 mutations on complex stability and functionality

To understand the functional role of the patient-derived *USP7* mutations, I transduced DLBCL cell lines with WT and four USP7 variants: Y243D, I660K, T730S, and Y1056H. Importantly, Y243D, I660K, and Y1056H mutations were found to be unstable upon treatment with cycloheximide, while T730S remained as stable as WT USP7, suggesting that the mutations significantly shorten the half-life of USP7 protein stability. Despite this instability in three USP7 variants, no significant changes in cell proliferation were observed.

The USP7 mutations that I have functionally assessed in this study (Y243D, I660K, T730S, and Y1056H) showed instability, particularly Y243D and I660K, as the pathogenicity score predicted. To see whether USP7 mutations were able to interact with other PRC1.1 subunits, I pulled these down and observed that the interaction with PCGF1 and BCOR was lost in Y243D and I660K mutated USP7.

In addition, there was indeed a disrupted interaction with other PRC1.1 subunits, such as BCOR and PCGF1, compromising complex stability (***Figure 25B***). When investigating the Y243D, I660K, and Y1056H mutations more closely, mutation Y243D introduces a change from a hydrophobic tyrosine (Y) to a negatively charged aspartic acid (D), which likely disrupts the hydrophobic interactions that are critical for maintaining the catalytic pocket and substrate binding in the UCH domain. In I660K, the replacement of isoleucine (I) with lysine (K) introduces a positive charge that most likely alters the electrostatic environment of the UBL domains, which could weaken binding affinity with BCOR and PCGF1, ultimately impairing complex assembly. As for the Y1056H mutation, which is in the C-terminal UBL domain, there is a replacement of tyrosine (Y) with histidine (H), introducing a positive charge. This region is essential for the proper catalytic function of USP7, as it folds back into the catalytical pocket. Alterations in this domain could destabilize the catalytic configuration and enzymatic activity of the complex. I have seen that binding with PCGF1 and BCOR is partially kept in one of the model systems used, indicating that this mutation might not fully compromise the PRC1.1 assembly. These mutations seen in *USP7* support the hypothesis that PRC1.1 has a tumor-suppressive role that gets disrupted in lymphomagenesis of DLBCL. The disruption becomes more apparent when USP7 is inhibited pharmacologically. In this case, a marked reduction in H2AK119ub levels suggests that the enzymatic function of PRC1.1 is significantly compromised without USP7's deubiquitinase activity. This raises the possibility that USP7 might be critical not just for the physical assembly of the complex, but for sustaining its catalytic activity. Without USP7, PRC1.1 seems unable to effectively maintain its repressive role through histone ubiquitination. It becomes clear that USP7 may not be simply a binding partner but could be essential for the proper orchestration of PRC1.1's repressive capabilities. The reduction in H2AK119ub upon USP7 inhibition suggests that the overall functionality of the repressive complex is lost[111]. This prompts a deeper investigation into the dual role of USP7 as both a structural component and a catalytic regulator within PRC1.1, and whether mutations and inhibition disrupt these roles in distinct ways.

4.8 Importance of structural predictions and PRC1.1 assembly

To further validate what had been observed experimentally, *in silico* models of the complex were developed to understand in a three-dimensional way where in the protein and the complex the mutations occurred and how they affect protein interaction. One of the main advancements of this study was, in fact, the generation of structural predictions for full-length BCOR and the full assembly of the PRC1.1 complex, specifically incorporating the subunits USP7, BCOR, KDM2B, and PCGF1. This assembly is particularly important because it allows for a deeper understanding of how mutations, particularly in USP7, disrupt the physical interactions within the complex, leading to the destabilization of PRC1.1 and therefore reduced functionality. The structural predictions revealed important insights into the binding interfaces between these proteins, which are critical for the proper function of the complex. For instance, by identifying the contact interfaces between USP7 and BCOR, it was possible to identify important binding sites between these two proteins and how mutations occurring in certain regions could disturb the stability of the complex.

For instance, the Y243D USP7 is located in the UCH domain, which is critical for interaction with BCOR, making it likely that this mutation disrupts complex stability. However, as *in silico* predictions suggest, the impact of such a mutation will heavily depend on the specific amino substitution (***Table 31***) and the resulting change in charge and protein conformation. The mutation I660K, despite not being directly in the specific USP7-BCOR predicted binding site, can also disturb the contact, as it is in the UBL1 domain, which is also important for protein-protein interactions, potentially affecting USP7's binding capability. By modeling these mutations, structural changes in USP7 and how these affect PRC1.1 assembly are crucial for delineating the molecular mechanisms by which PRC1.1 mutations contribute to lymphomagenesis of DLBCL. Specifically, if *USP7* mutations prevent it from interacting with BCOR or PCGF1, this could explain the loss of PRC1.1's repressive functions in DLBCL. Furthermore, this structural understanding opens new possibilities for therapeutic intervention by identifying specific residues that could be targeted to restore complex functionality or inhibit oncogenic pathways resulting from PRC1.1 disruption.

5 Outlook

Investigating PRCs, especially PRC1.1, is inherently challenging due to their nature as multiprotein assemblies and context dependencies. In addition, many subunits have roles beyond the PRC1.1 complex itself. Subunits like USP7 and BCOR are involved in various cellular contexts independently of PRC1.1, adding complexity to deciphering their specific contributions within the complex in the context of DLBCL. For instance, USP7 is also a regulator of p53[103]. This makes it difficult to distinguish whether observed effects result from PRC1.1 activity or independent functions of its components.

The fact that I can demonstrate a reduced half-life of USP7 and perturbed interactions by the mutations between these subunits in DLBCL is already a significant step. The finding that the PRC1.1 complex assembles when sufficient levels of its constituents are present and functions in this context, despite the overlapping, PRC1.1-independent roles of its components, represents a critical achievement. This underscores the complexity of Polycomb biology, where individual parts serve multiple, sometimes divergent functions. In this context, demonstrating that PRC1.1 subunits do assemble and are functionally relevant in DLBCL is a pivotal discovery, laying the foundation for deeper insights into their epigenetic regulation and potential therapeutic applications.

However, while this study has focused on identifying mutations and their basic functional implications, there is still much to be explored regarding how these mutations affect downstream gene regulation and the broader epigenetic landscape in DLBCL. The functional consequences of PRC1.1 mutations, particularly their impact on gene expression pathways and chromatin dynamics, remain underexplored.

5.1 Identifying PRC1.1-regulated loci in DLBCL

To address the stark contrast of having mechanistic data on how mutations in PRC1.1 subunits shorten the half-life of USP7 and the assembly of the PRC1.1 complex, but having no impact on the survival of cells, prompted me to perform CUT&Tag against the H2AK119Ub mark. Specifically, I used the model cells with PCGF1-induced H2AK119Ub levels and USP7 inhibition to perform CUT&Tag for the H2AK119Ub mark. In addition, I also submitted the cells for RNA-Seq. The combined data will allow me to identify the downstream transcriptional circuits modulated by the PRC1.1 complex. These experiments will be essential to confirm whether the genes that are *bona fide* PRC1.1 targets in GCB-type DLBCL. However, the data is not available in time for the submission deadline of this thesis.

Once the direct transcriptional targets of PRC1.1 have been identified, one additional line of investigation will be to elucidate the binding partners and chromatin localization of wild-type *versus* mutant PRC1.1 complexes, by performing additional CUT&Tag and mass spectrometry to reveal how mutations alter the functional dynamics of PRC1.1 and the binding partners in the chromatin context.

5.2 Performing functional studies on the involved pathways

One main limitation is the lack of a strong cellular phenotype. I could not show changes in the survival of cells despite having a strong chromatin change. Depending on the results in the CUT&Tag alternative pathways, other than survival, might need more attention. This also includes understanding the role of USP7 mutations in tumor development, by the generation of an *in vitro* lymphoma model and/or in a genetically engineered mouse model harboring *USP7* mutations identified in DLBCL patients.

5.3 Assessing the functional consequences for the PRC1.1-independent role of the USP7 mutations

As mentioned above, many of the PRC1.1 subunits have PRC1.1-independent functions. Most prominent is the role of USP7, also known as HASUP, for the regulation of p53-MDM2-MDM4 axis. To address this role, I will leverage the generated DLBCL cell lines with wildtype or USP7 mutations and perform global proteome assessment using time of flight mass spectrometry with and without USP7 inhibitor treatment.

6 References

1. Janeway Jr, C. A., Travers, P., Walport, M. & Shlomchik, M. J. *Immunobiology: The Immune System in Health and Disease*. (Garland Science, New York, 2001).
2. Murphy, K. & Weaver, C. *Janeway's Immunobiology*. (Garland Science, New York, 2016).
3. Abbas, A. K., Lichtman, A. H. & Pillai, S. *Basic Immunology: Functions and Disorders of the Immune System*. (Elsevier, Philadelphia, Pennsylvania, 2021).
4. Allen, C. D., Okada, T. & Cyster, J. G. *Germinal Center Organization and Cellular Dynamics Role of the GC in Antibody Responses*.
5. Gomes de Castro, M. A. *et al.* Differential organization of tonic and chronic B cell antigen receptors in the plasma membrane. *Nat Commun* **10**, 820 (2019).
6. Roco, J. A. *et al.* Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity* **51**, 337-350.e7 (2019).
7. Elgueta, R. *et al.* Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunological Reviews* vol. 229 152–172 Preprint at https://doi.org/10.1111/j.1600-065X.2009.00782.x (2009).
8. Kawabe, T. *et al.* The immune responses in CD40-deficient mice: Impaired immunoglobulin class switching and germinal center formation. *Immunity* **1**, 167–178 (1994).
9. Takemori, T., Kaji, T., Takahashi, Y., Shimoda, M. & Rajewsky, K. Generation of memory B cells inside and outside germinal centers. *Eur J Immunol* **44**, 1258–1264 (2014).
10. Taylor, J. J., Pape, K. A. & Jenkins, M. K. A germinal center–independent pathway generates unswitched memory B cells early in the primary response. *Journal of Experimental Medicine* **209**, 597–606 (2012).
11. Victora, G. D. *et al.* Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **143**, 592–605 (2010).
12. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* **15**, 137–148 (2015).
13. Muramatsu, M. *et al. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme*. *Cell* vol. 102 (2000).
14. Basso, K. & Dalla-Favera, R. Roles of BCL6 in normal and transformed germinal center B cells. *Immunol Rev* **247**, 172–183 (2012).
15. Allen, C. D. C. *et al.* Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat Immunol* **5**, 943–952 (2004).
16. Victora, G. D. & Nussenzweig, M. C. Germinal Centers. *Annu Rev Immunol* **30**, 429–457 (2012).
17. Palm, A.-K. E. & Henry, C. Remembrance of Things Past: Long-Term B Cell Memory After Infection and Vaccination. *Front Immunol* **10**, (2019).
18. Gonzalez, D. G. *et al.* Nonredundant Roles of IL-21 and IL-4 in the Phased Initiation of Germinal Center B Cells and Subsequent Self-Renewal Transitions. *The Journal of Immunology* **201**, 3569–3579 (2018).
19. Victora, G. D. & Nussenzweig, M. C. Germinal Centers. **30**, 57 (2024).
20. Calado, D. P. *et al.* The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat Immunol* **13**, 1092–1100 (2012).
21. Wishnie, A. J., Chwat-Edelstein, T., Attaway, M. & Vuong, B. Q. BCR Affinity Influences T-B Interactions and B Cell Development in Secondary Lymphoid Organs. *Front Immunol* **12**, (2021).
22. Shapiro-Shelef, M. *et al.* Blimp-1 Is Required for the Formation of Immunoglobulin Secreting Plasma Cells and Pre-Plasma Memory B Cells. *Immunity* **19**, 607–620 (2003).

23. Klein, U. *et al.* Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nat Immunol* **7**, 773–782 (2006).
24. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat Rev Immunol* **15**, 172–184 (2015).
25. Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**, 679–690 (2018).
26. Schmitz, R. *et al.* Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine* **378**, 1396–1407 (2018).
27. Wright, G. *et al. A Gene Expression-Based Method to Diagnose Clinically Distinct Subgroups of Diffuse Large B Cell Lymphoma*. vol. 100 www.pnas.orgcgidoi10 (2003).
28. Sehn, L. H. & Salles, G. Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine* **384**, 842–858 (2021).
29. Tilly, H. *et al.* Polatuzumab Vedotin in Previously Untreated Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine* **386**, 351–363 (2022).
30. Coiffier, B. *et al.* CHOP Chemotherapy plus Rituximab Compared with CHOP Alone in Elderly Patients with Diffuse Large-B-Cell Lymphoma. *New England Journal of Medicine* **346**, 235–242 (2002).
31. Sehn, L. H. & Gascoyne, R. D. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **125**, 22–32 (2015).
32. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
33. Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**, 1851–1861 (2005).
34. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* **43**, 830–837 (2011).
35. Lenz, G. *et al.* Stromal Gene Signatures in Large-B-Cell Lymphomas. *New England Journal of Medicine* **359**, 2313–2323 (2008).
36. Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88–92 (2010).
37. Ngo, V. N. *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
38. Wright, G. W. *et al.* A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell* **37**, 551-568.e14 (2020).
39. Béguelin, W. *et al.* EZH2 Is Required for Germinal Center Formation and Somatic EZH2 Mutations Promote Lymphoid Transformation. *Cancer Cell* **23**, 677–692 (2013).
40. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
41. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181–185 (2010).
42. Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117**, 2451–2459 (2011).
43. Donaldson-Collier, M. C. *et al.* EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. *Nat Genet* **51**, 517–528 (2019).
44. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
45. Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med* **21**, 1199–1208 (2015).
46. Zan, H. & Casali, P. Editorial: Epigenetics of B Cells and Antibody Responses. *Front Immunol* **6**, (2016).

47. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
48. Bird, A. DNA methylation patterns and epigenetic memory. *Genes and Development* vol. 16 6–21 Preprint at https://doi.org/10.1101/gad.947102 (2002).
49. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295–304 (2009).
50. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484–492 (2012).
51. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science (1979)* **293**, 1074–1080 (2001).
52. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011).
53. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).
54. Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570 (1978).
55. Jürgens, G. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316**, 153–155 (1985).
56. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *EMBO J* **4**, 3259–3264 (1985).
57. Shen, X. *et al.* EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Mol Cell* **32**, 491–502 (2008).
58. Sneeringer, C. J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. https://doi.org/10.1073/pnas.1012525107/-/DCSupplemental doi:10.1073/pnas.1012525107/-/DCSupplemental.
59. Béguelin, W. *et al.* EZH2 and BCL6 Cooperate to Assemble CBX8-BCOR Complex to Repress Bivalent Promoters, Mediate Germinal Center Formation and Lymphomagenesis. *Cancer Cell* **30**, 197–213 (2016).
60. Knutson, S. K. *et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Mol Cancer Ther* **13**, 842–854 (2014).
61. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in cancer. *Nat Med* **22**, 128–134 (2016).
62. Morschhauser, F. *et al.* Tazemetostat for patients with relapsed or refractory follicular lymphoma: an open-label, single-arm, multicentre, phase 2 trial. *Lancet Oncol* **21**, 1433–1442 (2020).
63. Erokhin, M. *et al.* Clinical Correlations of Polycomb Repressive Complex 2 in Different Tumor Types. *Cancers (Basel)* **13**, 3155 (2021).
64. Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722–726 (2010).
65. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
66. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
67. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol Cell* **45**, 344–356 (2012).
68. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell* **74**, 1020-1036.e8 (2019).
69. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol Cell* **45**, 344–356 (2012).
70. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147–1155 (2013).

71. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643–649 (2015).
72. Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P. & Brockdorff, N. Chromatin Sampling—An Emerging Perspective on Targeting Polycomb Repressor Proteins. *PLoS Genet* **9**, e1003717 (2013).
73. Kundu, S. *et al.* Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell* **65**, 432-446.e5 (2017).
74. Isshiki, Y. & Iwama, A. Emerging role of noncanonical polycomb repressive complexes in normal and malignant hematopoiesis. *Exp Hematol* **68**, 10–14 (2018).
75. Damm, F. *et al.* BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. *Blood* **122**, 3169–3177 (2013).
76. Astolfi, A. *et al.* BCOR Involvement in Cancer. *Epigenomics* **11**, 835–855 (2019).
77. Everett, R. D. A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. *EMBO J* **16**, 1519–1530 (1997).
78. Wang, Z. *et al.* USP7: Novel drug target in cancer therapy. *Frontiers in Pharmacology* vol. 10 Preprint at https://doi.org/10.3389/fphar.2019.00427 (2019).
79. Sánchez, C. *et al.* Proteomics Analysis of Ring1B/Rnf2 Interactors Identifies a Novel Complex with the Fbxl10/Jhdm1B Histone Demethylase and the Bcl6 Interacting Corepressor. *Molecular & Cellular Proteomics* **6**, 820–834 (2007).
80. Lecona, E., Narendra, V. & Reinberg, D. USP7 Cooperates with SCML2 To Regulate the Activity of PRC1. *Mol Cell Biol* **35**, 1157–1168 (2015).
81. Maat, H. *et al.* The USP7-TRIM27 axis mediates non-canonical PRC1.1 function and is a druggable target in leukemia. *iScience* **24**, (2021).
82. Qi, S.-M. *et al.* Targeting USP7-Mediated Deubiquitination of MDM2/MDMX-p53 Pathway for Cancer Therapy: Are We There Yet? *Front Cell Dev Biol* **8**, (2020).
83. Cartel, M. *et al.* Inhibition of ubiquitin-specific protease 7 sensitizes acute myeloid leukemia to chemotherapy. *Leukemia* **35**, 417–432 (2021).
84. Chauhan, D. *et al.* A Small Molecule Inhibitor of Ubiquitin-Specific Protease-7 Induces Apoptosis in Multiple Myeloma Cells and Overcomes Bortezomib Resistance. *Cancer Cell* **22**, 345–358 (2012).
85. Turnbull, A. P. *et al.* Molecular basis of USP7 inhibition by selective small-molecule inhibitors. *Nature* **550**, 481–486 (2017).
86. Wong, S. J. *et al.* Structure and Role of BCOR PUFD in Noncanonical PRC1 Assembly and Disease. *Biochemistry* **59**, 2718–2728 (2020).
87. Wong, S. J. *et al.* KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1. *Structure* **24**, 1795–1801 (2016).
88. Takano, J. *et al.* PCGF1-PRC1 links chromatin repression with DNA replication during hematopoietic cell lineage commitment. *Nat Commun* **13**, 7159 (2022).
89. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
90. Hatzi, K. *et al.* A Hybrid Mechanism of Action for BCL6 in B Cells Defined by Formation of Functionally Distinct Complexes at Enhancers and Promoters. *Cell Rep* **4**, 578–588 (2013).
91. Ghetu, A. F. *et al.* Structure of a BCOR Corepressor Peptide in Complex with the BCL6 BTB Domain Dimer. *Mol Cell* **29**, 384–391 (2008).
92. Schaefer, E. J. *et al.* BCOR and BCORL1 Mutations Drive Epigenetic Reprogramming and Oncogenic Signaling by Unlinking PRC1.1 from Target Genes. *Blood Cancer Discov* **3**, 116–135 (2022).
93. Grossmann, V. *et al.* Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* **118**, 6153–6163 (2011).

94. Tanaka, T. *et al.* Internal deletion of BCOR reveals a tumor suppressor function for BCOR in T lymphocyte malignancies. *Journal of Experimental Medicine* **214**, 2901–2913 (2017).
95. Tara, S. *et al.* Bcor insufficiency promotes initiation and progression of myelodysplastic syndrome. *Blood* **132**, 2470–2483 (2018).
96. Zhang, A. *et al.* BCOR Mutations in Acute Myeloid Leukemia: Clonal Evolution and Prognosis. *Blood* **136**, 4–4 (2020).
97. Terada, K. *et al.* Usefulness of *BCOR* gene mutation as a prognostic factor in acute myeloid leukemia with intermediate cytogenetic prognosis. *Genes Chromosomes Cancer* **57**, 401–408 (2018).
98. Li, X. *et al.* Dynamics of epigenetic regulator gene BCOR mutation and response predictive value for hypomethylating agents in patients with myelodysplastic syndrome. *Clin Epigenetics* **13**, (2021).
99. Sportoletti, P., Sorcini, D. & Falini, B. BCOR gene alterations in hematologic diseases. *Blood* **138**, 2455–2468 (2021).
100. Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, (2012).
101. He, J., Nguyen, A. T. & Zhang, Y. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **117**, 3869–3880 (2011).
102. Pozhidaeva, A. & Bezsonova, I. USP7: Structure, substrate specificity, and inhibition. *DNA Repair (Amst)* **76**, 30–39 (2019).
103. Sheng, Y. *et al.* Molecular recognition of p53 and MDM2 by USP7/HAUSP. *Nat Struct Mol Biol* **13**, 285–291 (2006).
104. de Bie, P., Zaaroor-Regev, D. & Ciechanover, A. Regulation of the Polycomb protein RING1B ubiquitination by USP7. *Biochem Biophys Res Commun* **400**, 389–395 (2010).
105. Brune, V. *et al.* Origin and pathogenesis of nodular lymphocyte–predominant Hodgkin lymphoma as revealed by global gene expression analysis. *J Exp Med* **205**, 2251–2268 (2008).
106. Basu, A. *et al.* An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules. *Cell* **154**, 1151–1161 (2013).
107. Seashore-Ludlow, B. *et al.* Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov* **5**, 1210–1223 (2015).
108. Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* **12**, 109–116 (2016).
109. Kim, R. Q., van Dijk, W. J. & Sixma, T. K. Structure of USP7 catalytic domain and three Ubl-domains reveals a connector α-helix with regulatory role. *J Struct Biol* **195**, 11–18 (2016).
110. Zhao, S. *et al.* Mutational landscape of uterine and ovarian carcinosarcomas implicates histone genes in epithelial–mesenchymal transition. *Proceedings of the National Academy of Sciences* **113**, 12238–12243 (2016).
111. Maat, H. *et al.* The USP7-TRIM27 axis mediates non-canonical PRC1.1 function and is a druggable target in leukemia. *iScience* **24**, 102435 (2021).
112. Faesen, A. C. *et al.* Mechanism of USP7/HAUSP Activation by Its C-Terminal Ubiquitin-like Domain and Allosteric Regulation by GMP-Synthetase. *Mol Cell* **44**, 147–159 (2011).
113. Scelfo, A. *et al.* Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Mol Cell* **74**, 1037-1052.e7 (2019).
114. Junco, S. E. *et al.* Structure of the Polycomb Group Protein PCGF1 in Complex with BCOR Reveals Basis for Binding Selectivity of PCGF Homologs. *Structure* **21**, 665–671 (2013).
115. Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494.e15 (2017).

116. Tanaskovic, N. *et al.* Polycomb group ring finger protein 6 suppresses Myc-induced lymphomagenesis. *Life Sci Alliance* **5**, (2022).
117. Parry, E. M. *et al.* Evolutionary history of transformation from chronic lymphocytic leukemia to Richter syndrome. *Nat Med* **29**, 158–169 (2023).