

**KRAB zinc finger proteins link heterochromatin  
maintenance to replicative stress and inflammation in  
Diffuse Large B Cell Lymphoma**

Présentée le 23 juin 2023

Faculté des sciences de la vie  
Laboratoire de virologie et génétique  
Programme doctoral en approches moléculaires du vivant

pour l'obtention du grade de Docteur ès Sciences

par

**Filipe Amândio**  
**BRANDÃO SANCHES VONG MARTINS**

Acceptée sur proposition du jury

Prof. E. Oricchio, présidente du jury  
Prof. D. Trono, Dr P. Turelli, directeurs de thèse  
Prof. K. Burns, rapporteuse  
Prof. O. Michielin, rapporteur  
Prof. D. Hanahan, rapporteur

**Science begins with counting. To understand a phenomenon, a scientist must first describe it; to describe it objectively, he must first measure it. If cancer medicine was to be transformed into a rigorous science, then cancer would need to be counted somehow-measured in some reliable, reproducible way.**

**- Siddhartha Mukherjee  
*The Emperor of All Maladies: A Biography of Cancer***

# Acknowledgments

I would like to thank:

Prof. Didier Trono for his mentorship, support, and enriching discussions about science and life.

Dr. Priscilla Turelli for her support, kindness, and constructive comments about this work.

Sandra Offner and Charlène Raclot for being the most gifted wet lab technicians, but also the most patient and friendly persons I had the chance to work with.

Séverine Reynard for the help she provided to me with my numerous requests all over these years and for being so friendly.

Marc Friedly for being my friend and for all the Covid testing adventures we went through.

Julien Duc for being my friend and for our discussions on movie classics and his help in improving my computational skills.

Evarist Planet for his work on this project and our discussions about life and cancer.

Joana Carlevaro-Fita for her essential contribution to this project, her patience, and her enthusiasm.

My fellow PhD students and the Post-Docs of the Laboratory of Virology and Genetics (LVG) for all their support and incentive.

Dr. Julien Pontis for sharing his vision and knowledge about epigenetics during all the years we shared the same desk and for being my friend.

Dr. Pierre-Yves Helleboid for contributing to my personal development in science and as a human being.

Roger Guindon, my childhood friend, for his support and all the memories we share.

My former colleagues at the hospitals of Fribourg and Lausanne for teaching me the basics of clinical practice, internal medicine, oncology, and hematology, and for all the memories we share.

Dr. Romain Forey for his essential contribution to this project and all the fruitful scientific discussions we had along the way.

Dr. Olga Rossopoff for her essential contribution to this project, the love we share for science, and, above all, for the one we share for each other.

My father, my mother, and the rest of my family for their support and for being present each time I needed it.

# Table of Contents

ACKNOWLEDGMENTS .....	3
TABLE OF FIGURES .....	5
ABSTRACT .....	6
KEYWORDS .....	6
RÉSUMÉ .....	7
MOTS CLÉS .....	7
INTRODUCTION.....	9
1. DIFFUSE LARGE B CELL LYMPHOMA, THE EMPEROR OF ALL LYMPHOMAS.....	9
1.1 From first descriptions to the discovery of its origins .....	9
1.2 From groundbreaking therapeutic successes to yet unsolved drawbacks.....	12
2. DLBCL, A NEOPLASM LIVING ON THE EDGE OF REPLICATIVE CATASTROPHE .....	16
2.2 Sources of replicative stress in cancer cells.....	18
2.3 Endogenous replicative stress, state of knowledge in DLBCL.....	22
2.4 Heterochromatin regions as replicative stress hotspots.....	23
2.5 Replicative stress, a source of cell-intrinsic inflammatory responses.....	25
3. KRAB ZINC FINGER PROTEINS AS MAINTAINERS OF CONSTITUTIVE HETEROCHROMATIN MEMORY.....	27
3.1 Transcription factors in cancer, an overview .....	27
3.2 KZFPs and their TE targets, a source of epigenetic diversity and cancer plasticity .....	29
AIMS OF THE THESIS.....	34
CONTRIBUTION .....	35
KRAB ZINC FINGER PROTEINS ZNF587/ZNF417 PROTECT LYMPHOMA CELLS FROM REPLICATIVE STRESS- INDUCED INFLAMMATION .....	36
INTRODUCTION.....	37
RESULTS .....	40
A DLBCL transcriptome-wide association study reveals KZFPs as predictors of poor prognosis .....	40
A KZFP gene cluster defines a DLBCL subset with cell-autonomous growth features .....	41
ZNF587/417 depletion impairs lymphoma cell growth and viability.....	43
ZNF587/417 depletion leads to a broad downregulation of KZFPs and alters the heterochromatin landscape of lymphoma cells.....	44
ZNF587/417 depletion triggers replicative stress in lymphoma cells.....	45
ZNF587/417 depletion leads to cell-intrinsic inflammation .....	46
DISCUSSION .....	50
MATERIAL AND METHODS .....	52
FIGURES AND LEGENDS .....	64
PERSPECTIVES .....	86
LIST OF ABBREVIATIONS.....	87
REFERENCES.....	90
CURRICULUM VITAE .....	111



# Table of Figures

Figure 1   A KZFP gene cluster as poor prognosis predictor in DLBCL.....	64
Figure 2   <b>KZFP<sup>High</sup> DLBCLs display cell-autonomous growth features.....</b>	66
Figure 3   <b>ZNF587/417 depletion impairs cell growth and viability.....</b>	68
Figure 4   <b>ZNF587/417 depletion alters the heterochromatin landscape. ....</b>	70
Figure 5   <b>ZNF587/417 depletion triggers replicative stress.....</b>	73
Figure 6   <b>ZNF587/417 depletion leads to cell-intrinsic inflammation. ....</b>	76
Figure 7   <b>ZNF587/417 depletion modifies the antigenic profile of lymphoma cells. ....</b>	77
Figure 8   <b>Figure S1 related to Figure 2: A module of KZFP genes in DLBCLs with cell-autonomous growth features. ....</b>	78
Figure 9   <b>Figure S2 related to Figure 3: ZNF587/417 depletion impairs lymphoma cell growth and viability. ....</b>	79
Figure 10   <b>Figure S3 related to Figure 4: ZNF587/417 depletion alters the heterochromatin landscape of lymphoma cells. ....</b>	81
Figure 11   <b>Figure S4 related to Figure 5: ZNF587/417 depletion triggers replicative stress in lymphoma cells.....</b>	82
Figure 12   <b>Figure S5 related to Figure 6: ZNF587/417 depletion leads to cell-intrinsic inflammation.....</b>	83

# Abstract

Genomic instability enhances cancer progression by favoring clonal diversity, yet uncontrolled replicative stress can lead to mitotic catastrophe and inflammatory responses promoting immune rejection. KRAB-containing zinc finger proteins (KZFPs) are epigenetic modulators, which for many control heterochromatin at transposable element (TE)-embedded regulatory sequences. We identified a cluster of 18 KZFPs associated with poor prognosis in diffuse large B cell lymphoma (DLBCL). We found their upregulation to correlate with increased copy number alterations and suppression of immune responses in tumor samples. Upon depleting two that target evolutionarily recent TEs, the primate-specific ZNF587 and ZNF417 paralogs, the proliferation of DLBCL cell lines was drastically impaired and replicative stress abruptly induced with marked alterations of the chromatin landscape and multiplication of DNA replication origins. Furthermore, *ZNF587/417* knockdown upregulated interferon/inflammatory-related genes through activation of the cGAS-STING DNA sensing pathway, augmented the susceptibility of tumor cells to macrophage-mediated phagocytosis and modified their immunogenicity through an increased surface expression of HLA-I and reshuffling of their immunopeptidome. ZNF587 and ZNF417 are thus pro-oncogenic factors allowing for higher degrees of genetic instability through attenuation of replicative stress and secondary inflammation, an influence that likely facilitates the clonal expansion, diversification, and immune evasion of cancer cells.

## Keywords

Diffuse Large B Cell Lymphoma (DLBCL), KRAB-containing zinc finger proteins (KZFPs), ZNF587, ZNF417, transposable elements (TEs), endogenous retroviruses (ERVs), heterochromatin, replication stress, inflammation, immunogenicity.

# Résumé

L'instabilité génomique promeut la progression du cancer en favorisant la diversité clonale, cependant, un stress réplicatif incontrôlé peut également entraîner des catastrophes mitotiques et des réponses inflammatoires favorisant le rejet immunitaire des cellules cancéreuses. Les protéines KRAB à doigt de zinc (KZFPs) sont des facteurs de transcription qui, pour beaucoup, contrôlent l'hétérochromatine au niveau des éléments transposables (TE). Nous avons identifié un groupe de 18 KZFPs associés à un mauvais pronostic dans le lymphome diffus à grandes cellules B (DLBCL). Nous avons constaté que leur régulation à la hausse était corrélée à une plus importante altération du nombre de copies de gènes, ainsi qu'à une diminution de l'infiltrat immun au sein de ces lymphomes. La déplétion de deux KZFPs ciblant des TE restreints sur le plan de l'évolution aux primates, nommément les paralogues ZNF587 et ZNF417, provoqua une diminution marquée de la prolifération des lignées cellulaires lymphomateuses testées. Le mécanisme sous-jacent étant lié à un stress réplicatif accentué, brusquement induit, associé à des altérations marquées du paysage chromatinien et à la multiplication d'origines de réplication de l'ADN. De plus, la diminution de ZNF587/417 déclencha la hausse de gènes liés à la réponse interféron/inflammatoire via l'activation de la voie cGAS-STING suite au relargage de fragments d'ADN nucléaire dans le cytosol. Cette inflammation endogène a augmenté la sensibilité des cellules lymphomateuses à la phagocytose médiée par les macrophages, et mené à une augmentation de leur immunogénicité par une expression de surface accrue des complexes HLA-I et à un remaniement de leur immunopeptidome. ZNF587 et ZNF417 sont donc des facteurs pro-oncogéniques permettant des degrés plus élevés d'instabilité génomique grâce à l'atténuation du stress réplicatif et de l'inflammation qui en découle, une influence qui facilite probablement l'expansion clonale, la diversification et l'évasion immunitaire des cellules lymphomateuses.

## Mots clés

Lymphome diffus à grandes cellules B (DLBCL), protéines KRAB à doigts de zinc (KZFP), ZNF587, ZNF417, éléments transposables (TE), rétrovirus endogènes (ERV), hétérochromatine, stress réplicatif, inflammation, immunogénicité.



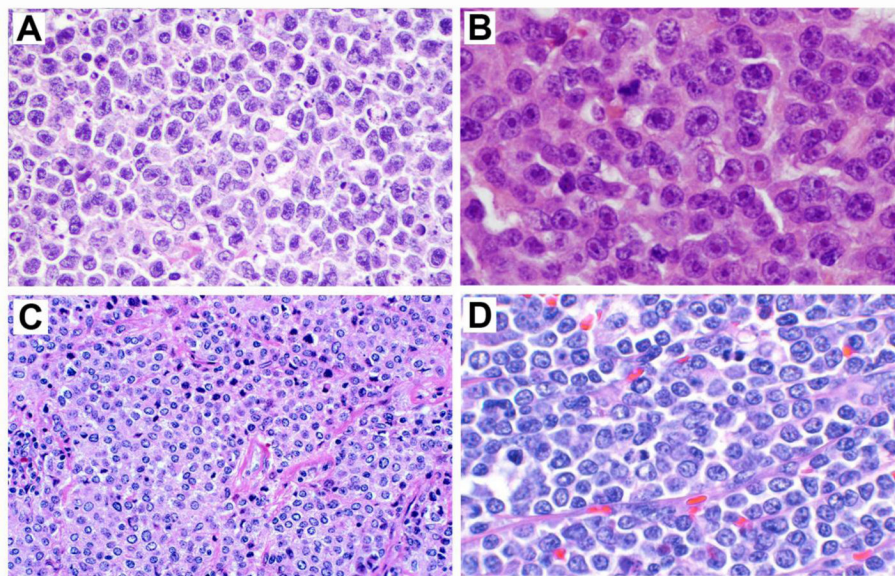
# Introduction

## 1. Diffuse Large B Cell Lymphoma, the emperor of all lymphomas

### 1.1 From first descriptions to the discovery of its origins

When the British pathologist Thomas Hodgkin first described the disease that will later carry his name in 1832, he was stunned by the enlarged spleen and swollen lymph nodes located in the neck, axilla, and groins, but also by their close proximity to large blood vessels in affected patients<sup>1</sup>. He was actually observing cancer evolving in the lymphatic system. During the following decades, other pathologists continuously reported cases with strikingly similar presentations. Nevertheless, it soon appeared that those rather stereotypic presentations were only present in a fraction of patients succumbing to this malignant form of nodal enlargement, which was named “lymphoma” by the Prussian surgeon Theodor Billroth in 1871<sup>2</sup>. Optical microscopy allowed the identification of the cells responsible for Hodgkin’s disease by Dorothy Reed and Carl Sternberg in 1898<sup>3,4</sup>. Nevertheless, at that time pathologists were still not able to classify the remaining bulk of more heterogenous types of lymphomas. It was only in the late 1960s that the pathologist Henry Rapoport was able to propose the first reliable description of what is now called the non-Hodgkin lymphoma (NHL), presenting either as diffuse, nodal, or histiocytic types of monotonous cell infiltrates<sup>5</sup>. During the same period, it became also clear that NHL was at least thrice more frequent than Hodgkin lymphoma (HL)<sup>6</sup>. A more functional classification of lymphoid neoplasms was proposed by Lukes and Collins in the mid 1970s<sup>7</sup>, as they combined advances in the immunology field which was now able to determine with confidence the B or T cell of origin of cells, with a deeper description of the site of origin in the lymph nodes, and additional morphological features of the nucleus and size of malignant cells, allowing the identification of approximatively 30 different NHL subtypes. These initial histology-based classification systems allowed the first epidemiological surveys of lymphoma incidence which soon revealed large B cell lymphomas with diffuse infiltration patterns, to be the most frequent type of lymphoma, accounting for more than a third of all NHL cases. Although these classifications constituted the initial steps in dissecting these complex lymphoid malignancies and asserting their diagnoses, they were still not reunited into entities with clinical relevance, as their diagnoses were still subjected to some subjective criteria hindering broad acceptance by medical communities. It required an initial trans-atlantic

effort of lymphoma study groups, that culminated in a revised european-american lymphoma (REAL) classification in the mid-1990s<sup>8</sup>, to obtain a first consensual view of lymphoma clinico-biological entities and respective diagnostic criteria. Diffuse large B cell lymphoma (DLBCL) became, at that time, a unified entity, apart from its Burkitt (BL) and primary mediastinal B cell lymphoma (PMBCL) relatives. Perhaps, one of the most valuable contributions of the REAL classification was the inclusion of immunophenotype and chromosomal abnormalities in disease subtype characterization. However, no correlation was found between immunophenotype and DLBCL histology, which was so far described as a diffuse cell infiltrate composed of a variable mixture of centroblast- (characterized by a large, noncleaved nucleus, and peripherally-located, often multiple, nucleoli) and immunoblast-like cells (typically larger than centroblasts, these B cells present a larger centrally-located nucleolus). One-third of cases were reported to harbor *BCL2* gene rearrangements and another smaller proportion carried rearrangements of the *c-MYC* oncogene.



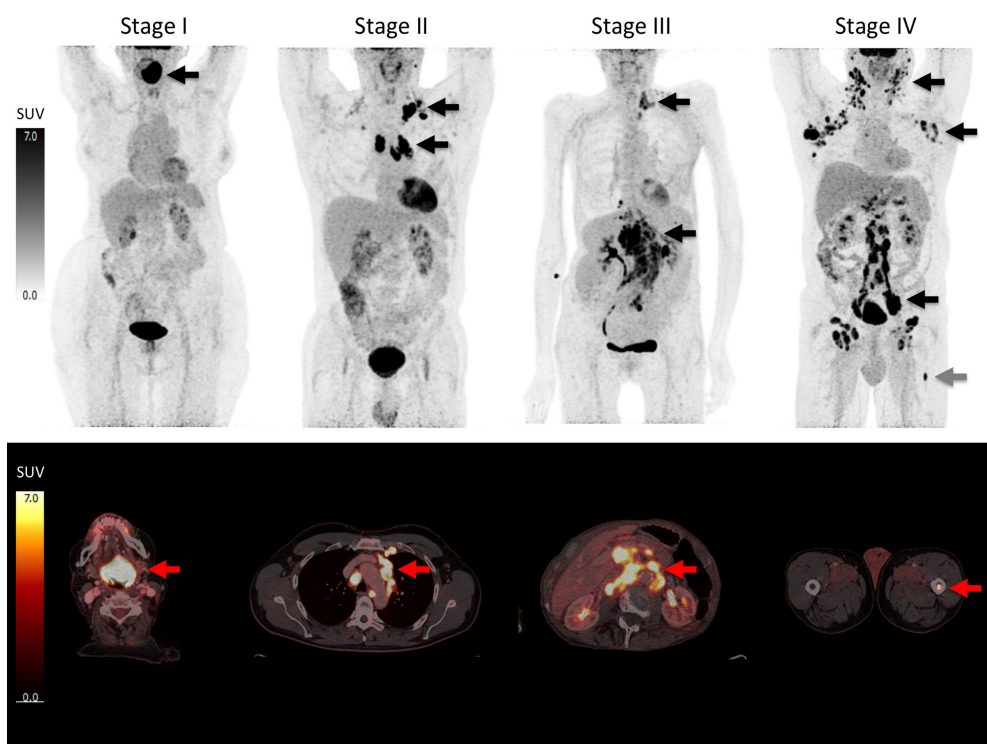
**Figure 1 | DLBCL subtypes.** (A) DLBCL, centroblastic variant. (B) DLBCL, immunoblastic variant. (C) PMBCL. (D) BL. Hematoxylin and eosin stainings.<sup>9</sup>

In spite of these findings, it was concluded that further subclassifying DLBCL was difficult with the methods available and of no practical utility at that time. Nevertheless, the disparities in treatment outcomes were strongly suggesting that DLBCL aggressivity was not solely determined by clinical and histological features. In the early 2000s, the advent of gene expression profiling using DNA microarrays allowed the distinction of two major DLBCL subtypes, one expressing genes characteristic of germinal center B cells (GCB) and another

presenting a gene expression profile similar to one seen during the *in vitro* activation of peripheral blood B-cells (ABC)<sup>10</sup>. The latter was found to be associated with worst outcomes and probably issued from a later stage of B cell differentiation closely related to one of the early plasmablasts, an intermediate stage along the way of germinal center B cell differentiation process towards the terminally differentiated plasma cell<sup>11</sup>. This discovery further supported the postulate of Lukes and Collins 30 years earlier, who claimed that lymphomas developed from a “block” in lymphocyte differentiation. Further genomic studies fueled this theory by showing enrichment of mutations in immunoglobulin genes, suggesting ongoing somatic hypermutation (SHM) in GCB but not in ABC, a core process in antibody maturation<sup>12,13,14</sup>. Furthermore, GCB-DLBCLs are also enriched in mutations affecting genes central to GC formation such as EZH2<sup>15</sup> and actors of the Gα13 pathway<sup>16</sup>. Mirroring GCB-DLBCLs, ABC-DLBCLs are enriched in mutations affecting genes involved in the terminal B cell differentiation process and/or leading to NF-κB constitutive activation, therefore mimicking constant B cell receptor (BCR) signaling<sup>17</sup>. Since the early 2000s, the efforts for constant lymphoma care and diagnostic improvement have been coordinated and continuously updated by the World Health Organization (WHO) which in its latest release recognizes more than 500 types of NHL and at least 4 HL subtypes<sup>18</sup>.

## 1.2 From groundbreaking therapeutic successes to yet unsolved drawbacks

Aggressive mature B cell lymphomas are perhaps one of the best examples of highly proliferative and exquisitely sensitive to DNA damaging agents' type of neoplasms. B cell lymphomas most often initiate in secondary lymphoid organs and tend to spread from one adjacent lymphoid organ to the next through the lymphatic vessels. The direct invasion of neighboring, or distant organs (i.e., so-called extra-nodal extension) through the bloodstream, is usually characteristic of its most aggressive forms.



**Figure 2 | The Ann Arbor staging classification.** Stage I = involvement confined to a single lymph node region or extranodal site; Stage II = involvement of more than one lymph node region on one side of the diaphragm; Stage III = lymph node involvement on both sides of the diaphragm; Stage IV = diffuse/disseminated extranodal involvement. SUV, Standardized Uptake Value. The letter E designating contiguous extranodal involvement can be added to stages I-III; The letter B designating the occurrence of systemic symptoms (encompassing unexplained fever, significant weight loss, and night sweats) can be added to all stages<sup>19</sup>.



First historical observations highlighting the sensitivity of lymphoma to DNA damage-inducing therapies surged even before the discovery of DNA itself, as the Swiss radiologist René Gilbert showed that HL was responsive to radiation therapy in the early 1930s<sup>20</sup>. Moreover, the clinical development of nitrogen mustard derivatives in the early 1950s allowed clinicians to induce first remissions in patients with lymphomas not amenable to extended field radiation therapy due to widespread diseases<sup>21</sup>. However, the remissions achieved with these alkylating agents were in their vast majority short-lived, until the establishment of first-generation protocols using chemotherapeutic combinations. The first 20 years of chemotherapy existence have been centered on developing, and chemically improving, the formulation of these new classes of cytostatic drugs, and learning to manage their toxicity. The latter preferentially affect organs subjected to high physiological turnovers such as the bone marrow, the digestive tube, and the teguments, thus requiring high-quality drug handling procedures by specialized teams and fine-tuned administration protocols.

First developed during the 1970s, a regimen associating an improved nitrogen mustard (cyclophosphamide), an anthracycline antibiotic (hydroxydaunorubicin), a mitotic spindle poison (vincristine, also known as vincristine), and high-dose steroids (prednisolone) imposed itself in the early 1990s as the standard treatment of aggressive mature B cell lymphomas, after being compared to other regimens in a milestone phase III trial<sup>22</sup>. More specifically, the discovery of the antitumor activity of the first anthracycline antibiotics extracted from strains of *Streptomyces peucetius* during the 1960s led to the synthesis of the molecule hydroxydaunomycin (i.e. doxorubicin)<sup>23</sup>, an inhibitor of the topoisomerase II enzyme involved in the release of supercoils generated by the DNA replication machinery, through the genesis of transient double-strand breaks (DSBs)<sup>24</sup>. This agent proved to be extremely effective against lymphoid neoplasms and is believed to be the core component of the CHOP regimen. This regimen achieved cure rates surrounding 30-40% which was a major breakthrough in this previously deadly disease<sup>25</sup>. Nevertheless, a cure was still not the rule, and improving CHOP has been the focus of active research efforts. Niels K. Jerne, Georges J.F. Köhler, and César Milstein received the Nobel Prize of Physiology and Medicine in 1984 for the discovery of the hybridoma technique, allowing the production of monoclonal antibodies (mAbs) through the somatic hybridization of myeloma cells with immunized B cells<sup>26</sup>. It allowed Lee M. Nadler and colleagues, at the Dana Farber Institute, to produce enough titers of an antibody directed against the lymphoma cells of a patient with an aggressive B cell lymphoma<sup>27</sup>. After ensuring that this antibody was not reactive against normal tissue samples, they tested it in a 54-year-old patient suffering from a DLBCL refractory to chemotherapy who experienced a transient

decrease in blood-circulating lymphoma cells and no side effects, thus proving the activity and feasibility of serotherapy in cancer treatment. It was subsequently discovered that the antigen targeted by their antibody was a transmembrane glycoprotein named CD20<sup>28,29</sup>. This antigen appeared to be specific to normal and malignant B cells, but started to be expressed at the pre-B cell stage of early B cell development and not at the pro-B cell earlier stage, thus preserving B cell precursors from anti-CD20 mAbs. The first generation of anti-CD20 mAbs was still poorly active as it only triggered complement activation and not yet antibody-dependent cellular cytotoxicity (ADCC). In 1990, the pharmaceutical company IDEC isolated a hybridoma able to produce a CD20-direct mAb with ADCC capacity, as it was able to cap at B cell surface upon engagement of its target, thus allowing Natural Killer (NK) cell recruitment<sup>30,31</sup>. The first multicenter trial published by McLaughlin et al.<sup>32</sup> testing this now-called rituximab antibody on 166 patients suffering from relapsing follicular lymphoma demonstrated tremendous activity with nearly half of the patients responding to therapy, leading to its rapid approval by the US Food and Drug Administration (FDA)<sup>33</sup>. When tested in combination to CHOP in 2001, the R-CHOP regimen achieved long-lasting complete remissions (CR) in more than 60% of patients, thus nearly doubling the curability of DLBCL<sup>34</sup>. A closer look at the patterns of treatment failure identified two faïths among the 40% of patients failing R-CHOP treatment, half presented a relapse, usually in the first one to two years following a virtual disappearance of their disease, and another half were not sensitive to it, presenting lymphomas that would not or only partially regress (a so-called refractory disease state)<sup>35</sup>. Therefore, predicting treatment-patient outcomes was a main concern and led to the identification of DLBCL prognostic biomarkers in the era of CHOP.

In the early 1980s, an international collaborative effort culminated in the elaboration of a prognostic model named IPI (standing for International Prognostic Index) based on the risk factors identified over 2'031 patients with DLBCL treated with CHOP across 15 different countries<sup>36</sup>. This model based on age, serum lactate dehydrogenase (LDH) levels, tumor stage, and the number of extranodal sites was able to reproducibly identify patient risk groups, and its updated version in the R-CHOP era is still routinely used in the clinic<sup>37</sup>. Although informative and useful for clinical trial design, high-risk patients do not clearly benefit from more intensive poly-chemotherapy regimens despite numerous studies. This is linked to the fact that treatment-related toxicity of poly-chemotherapy regimens seems to blur their survival benefit in the first-line setting, with until recently, high-dose regimens followed by autologous stem cell transplant (ASCT) support being reserved for selected patients presenting refractory/relapsing diseases. The detection of early signs of treatment failure to better select

the patients that could benefit from treatment intensification is under study and relies at present time on Positron Emission Tomography – Computed Tomography (PET-CT) by measuring the residual metabolic activity of lymphomas at intermediate stages during the course of R-CHOP treatment cycles<sup>38</sup>.

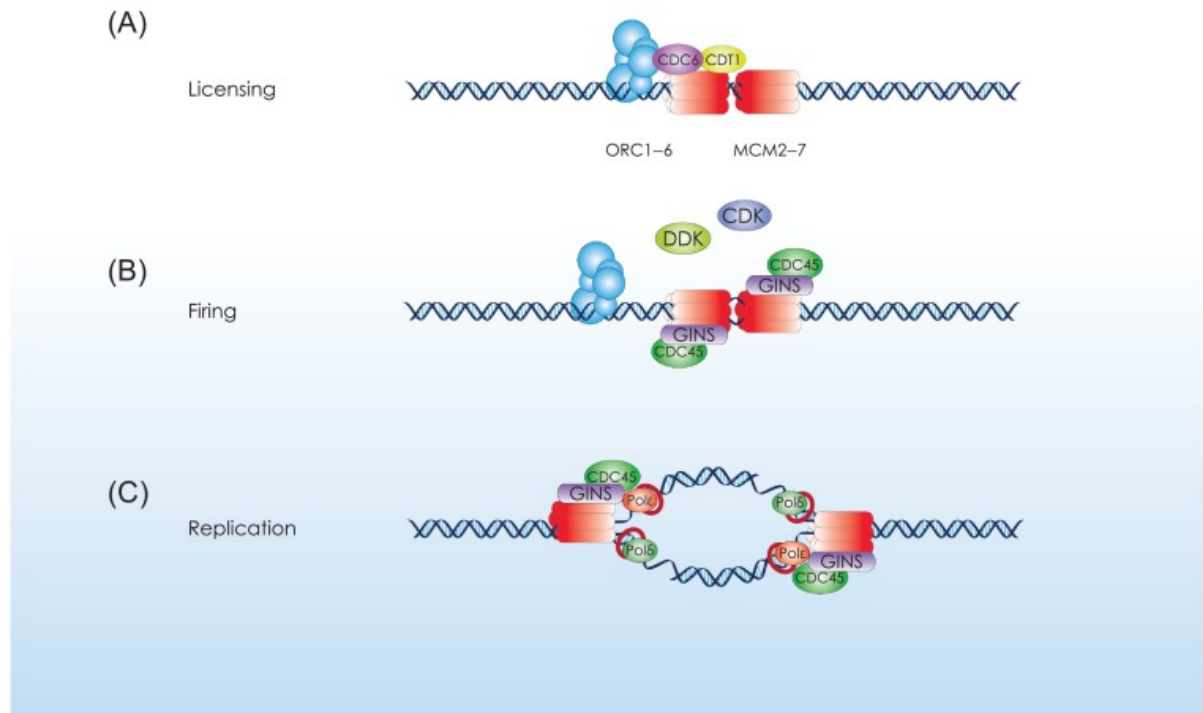
In 1998, Brian Drucker published the first clinical trial demonstrating the safety and efficacy of a small molecule targeting specifically the ATP pocket of a fusion protein (BCR-ABL) issued from a recurrent chromosomal translocation t(9;22) in Chronic Myelogenous Leukemia (CML)<sup>39</sup>. The revolution of target therapy had started and DLBCL would not be spared. Strong of the knowledge acquired from the molecular characterization of DLBCL, small molecules targeting kinases essential for lymphoma cell survival were developed. In 2010, a small molecule inhibitor of Bruton's tyrosine kinase, a key enzyme in the BCR signaling pathways was developed and demonstrated to be active in specific molecular ABC-DLBCL subtypes<sup>40</sup>. Moreover, the list of small molecules with clinical activity is in full expansion and encompasses inhibitors of the PI3K-AKT pathway<sup>41</sup> along with inhibitors of BCL2<sup>42</sup>, an often-upregulated anti-apoptotic protein preventing caspase activation<sup>43</sup>. These new compounds are currently under clinical investigation in randomized control trials but so far, they have not been able to improve the efficacy of R-CHOP when used in combination. Nevertheless, they are now part of the therapeutic arsenal to treat relapsed/refractory (R/R) DLBCL.

The most recent revolution in the treatment of DLBCL comes from the recent advances in gene editing allowing the engineering of the T cell receptor (TCR). Chimeric Antigen Receptor (CAR) T cells are produced from patients' own T cells to avoid rejection or graft-versus-host reactions<sup>44,45,46</sup>. T cells obtained through leukapheresis are subsequently genetically engineered to express a modified TCR recognizing the B cell-specific CD19 surface antigen, before being amplified and reinfused to the patient. CAR T cells have been initially used for the treatment of relapsing B acute lymphoblastic leukemia (B-ALL) in the pediatric setting since 2013<sup>47</sup>. Its approval for the treatment of DLBCL came in 2017 when axicabtagene ciloleucel (Yescarta)<sup>48</sup> and tisagenlecleucel (Kymriah)<sup>49</sup> products have been FDA-approved for the treatment of R/R DLBCL, after showing a 2-year progression-free survival exceeding 60%. It is estimated that at least half of R/R patients will achieve long-term remission with these recent advances in DLBCL therapy. Additional CAR T cells targeting CD20 and numerous other surface antigens in other hematological malignancies are currently in clinical development and will soon arrive to the clinic<sup>50</sup>. Although the war against DLBCL seems to come to an end, its incidence is increasing worldwide<sup>51</sup>, with 2-3 patients out of 10 still succumb to their disease.

## **2. DLBCL, a neoplasm living on the edge of replicative catastrophe**

### **2.1 Genomic instability and DNA replication**

Genomic instability designates the propensity of cancer cells to accumulate genetic alterations at a higher frequency than in healthy tissues. This characteristic is a main driver of cancer progression, as it promotes the emergence of cancer subclones harboring enough genetic diversity to survive the bottleneck selection of exogenous or endogenous selective pressures. This allows cancer to copy and adapt to various cytotoxic, genotoxic, and/or microenvironmental challenges, thus paving the way to treatment resistance, relapse, and disease progression. It can result from errors occurring during DNA replication, a process occurring before each cell division. Although taking place during the S phase of the cell cycle, DNA replication starts with the assembly of pre-replicative complexes (pre-RCs) during the entire G1 phase of the cell cycle, an event known as origin licensing<sup>52</sup>. Pre-RCs are composed of the Origin Recognition Complex (ORC) which binds DNA at replication origins, CDT1, CDC6, and the MCM proteins 2 to 7 constituting the hexameric DNA helicase<sup>53</sup>. The maturation of these complexes is under the control of cyclin-dependent kinases (CDKs) which participate to prevent further loading of pre-RCs upon the G1/S transition and permit the binding of CDC45 and the GINS complex onto pre-RCs, thereby increasing their enzymatic activity<sup>54</sup>. Further loading of the DNA polymerases pol $\epsilon$  and pol $\delta$  is the final step of the maturation of pre-RCs into replisomes<sup>55,56</sup>. The second phase of DNA replication is known as origin firing, a process in which replisomes start to unwind DNA and progress bidirectionally as replication forks, conducting their DNA synthesis activity<sup>57</sup>.



**Figure 3 | DNA replication<sup>58</sup>.** (A) **Origin licensing:** loading of the minichromosome maintenance complex 2–7 (MCM2–7) onto replication origins during the G1 phase of the cell cycle. Two MCM2–7 complexes are loaded at each origin through the activities of the six-subunit origin recognition complexes 1–6 (ORC1–6), the cell division cycle 10-dependent transcript 1 (CDT1), and the cell division cycle 6 (CDC6) proteins to form the pre-RC. (B) **Origin firing:** activation of the MCM2–7 helicases by cyclin-dependent (CDK) and DBF4-dependent kinases (DDK) and association of CDC45 and GINS during the S phase. (C) **Replication:** bidirectional replication fork progression, through the activity of the CDC45–MCM2–7–GINS complex coordinated by pol $\delta$  and pol $\epsilon$  DNA polymerases and elongation factors<sup>58</sup>.

The genome-wide firing of replication origins is tightly regulated spatially and temporally<sup>59</sup> as it has to (1) ensure a complete duplication of the genome by avoiding too large gaps between active origins, (2) avoid excessive consumption of limiting factors, (3) prevent collisions with the RNA transcription machinery. Therefore, only a subset of licensed origins become active, in a sequential and coordinated manner, in order to fulfill these requirements<sup>60</sup>. Conducting a proper replication program in the presence of short cell cycling<sup>61,62</sup>, intercurrent mutagenic events<sup>63</sup>, and exacerbated oncogene-induced transcriptional activity<sup>64</sup> is a challenge that

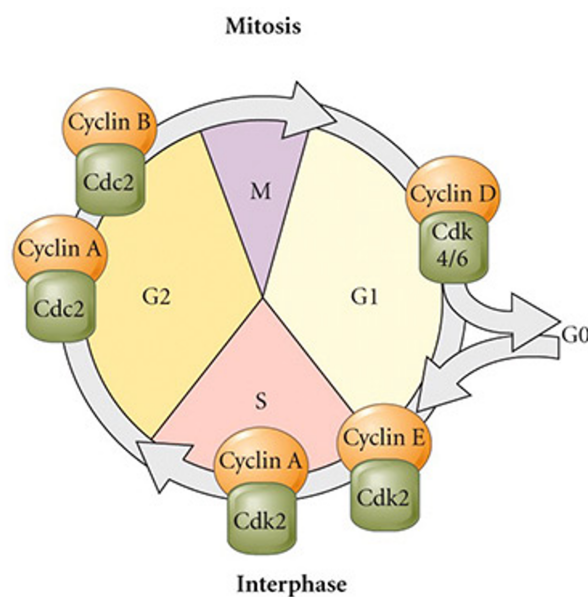
DLBCL cells have to face which renders them prone to replicative stress (RS). A type of genomic stress characterized by the stalling and potential collapse of replication forks leading to genomic instability<sup>65</sup>. Replication fork stalling provokes an uncoupling of the DNA polymerase and MCM helicase complex which generates single-strand DNA (ssDNA) segments. Aiming at preserving genome integrity, these ssDNA segments will be coated by the RPA protein allowing the recruitment of the ATR kinase<sup>66,67</sup>. ATR signaling delays cell cycle progression, stabilizes replication forks, and promotes their resumption. However, these protective mechanisms can be overrun, ultimately leading to fork collapse, activation of nucleases, and double-strand DNA breaks (DSBs)<sup>68</sup>.

## **2.2 Sources of replicative stress in cancer cells**

Causes impairing replication fork progression are numerous and may arise from endogenous or exogenous sources. Drug-induced replication stress is in fact harnessed for cancer therapy since the discovery of chemotherapy in the 1950s. Chemotherapy agents such as platin salts or alkylating mustard gas derivatives generate DNA cross-links which impair replication fork progression<sup>69</sup>. Additionally, drugs affecting the enzymatic activity of proteins associated with the replication machinery, such as topoisomerase inhibitors<sup>70</sup>, or causing the depletion of the nucleotide pool used by polymerase subunits for DNA synthesis, such as hydroxyurea<sup>71</sup>, are widely used in the clinic. Perhaps, radiation therapy is, through its genotoxic effects, the oldest example of the use of RS as a cancer treatment.

Endogenous sources of RS can be divided into three categories: the ones that generate obstacles to replication progression, the ones that lead to limiting factor shortage, and the ones that affect directly the replication program. These categories are not mutually exclusive as they might impact each other or even be causally related. Understanding the mechanisms underlying endogenous RS is of the utmost importance, as it could contribute to revealing weaknesses or therapeutic targets allowing more efficient and less toxic ways to generate RS in cancer cells. Moreover, evidence indicates that RS is one of the core initiating events linking oncogene activity with neoplastic transformation<sup>72</sup>. Disease models issued from cancer predisposition syndromes, in which a mutation in a critical gene, usually implicated in DNA repair, is already present and awaits a second mutation in the other allele to initiate cancer nourished the hypothesis that similar events were also at the origin of sporadic cancers. Moreover, the increased rate of mutations in cancer cells nourished the idea of an amplifying process in which mutations conferring proliferative advantages are initially selected, and clonal evolution to be

mostly driven by this mutator phenotype hypothesis<sup>73,74</sup>. However, the genome-wide characterization of hundreds of tumors caused research groups to question this hypothesis, as mutations in DNA repair genes were present in only a third of sporadic tumors<sup>75</sup>. Moreover, upstream surveillance mechanisms that coordinate and monitor DNA repair and cell cycle progression, which are exemplified by mutations in TP53 and CDKN2A (p16) are present in only 50% and 25% of tumors, respectively, with their frequency increasing in more advanced tumors, thus suggesting their role as sole initiating events to be unlikely. The recurrence of genetic abnormalities amongst some genes involved in growth signaling, together with their capacity to trigger genomic instability and neoplastic transformation upon overexpression of their wild-type forms, gave rise to a new paradigm, in which genetic or epigenetic alterations triggering the inappropriate hyperactivation of cell growth could be a major path of cancer initiation and progression.



**Figure 3 | Cell cycle.** Schematic view of the cell cycle of somatic cells (extracted from *Developmental Biology*, 9<sup>th</sup> edition, Figure 5.1 (Part1))<sup>76</sup>. Mitoses (M) are followed by an interphase stage subdivided into G1, S (synthesis), and G2 phases. The cyclins (orange) and their respective kinases (green) are shown at their phase of cell cycle regulation (known as checkpoints).

The cell cycle is made of phase transitions coordinated by the post-translational modification of CDKs, as their activity depends on their interaction with transiently expressed proteins known as cyclins. Each phase completion is under the control of a given checkpoint that ensures genome integrity before its transition to the next step. Evidence suggesting that disrupting the replication program, or its timing, could promote genomic instability has been shown through the overexpression of several main oncogenes, such as Ras<sup>77</sup>, cyclin E<sup>78</sup>, MDM2<sup>79</sup>, MYC<sup>80</sup>, or even viral oncogenes such as E6/E7 HPV-related proteins<sup>81</sup>, a phenomenon known as oncogene-induced RS.

How oncogenes trigger RS is still a matter of debate as it appears to be multifactorial. The global deregulation of origin licensing and firing appears to be at stake in the case of oncogenes involved in cell cycle progression. Defects in origin licensing as well as excessive origin firing have been implicated as mechanisms leading to RS in the case of cyclin E overexpression in epidermoid carcinoma<sup>82</sup> and osteosarcoma cell line models<sup>78</sup>, respectively. Therefore, the mechanisms underlying cyclin E-induced RS appeared to be dependent on cellular context. Further studies on CDKs showed that increasing their activity promoted RS by increasing origin firing to an extent capable of depleting intracellular pools of deoxynucleotides<sup>83</sup>. Replenishing the pool of deoxynucleotides was able to rescue this phenotype and indicated that activation of their synthesis is crucial to avoid RS in the setting of uncoordinated entry into the S phase. In addition to nucleotide availability, other limiting factors contributing to RS in the context of excessive origin firing have been discovered, such as RPA. Increased CDK activity using inhibitors of WEE1 (an inhibitor of CDKs) and impaired replication fork progression using hydroxyurea (an inhibitor of the ribonucleotide reductase) triggered the depletion of RPA stocks leading to fork collapse and DSBs<sup>84</sup>. Furthermore, the genome appears to be not equally sensitive to DNA damage induced by endogenous levels of RS.

In 1979, Sutherland and colleagues reported the influence of cell culture conditions on the frequency of chromosomal breaks/gaps at specific sites<sup>85</sup>. These sites were designated as common fragile sites (CFS), as they occurred recurrently in a given cell type under similar cell culture conditions. The discovery of the DNA polymerase and its inhibitory drug aphidicolin allowed the broad characterization of CFS across cell types<sup>86</sup>. CFS may span along megabases but appeared to be preferentially located in AT-rich regions, associated with long genes, heterochromatin, and sites with incompletely replicated DNA (i.e., late replication sites). Inhibition of transcription elongation in cyclin E or RAS-overexpressing cells reduced RS and DNA damage, thus indicating that conflicts between transcription and replication could account for the link between long genes and CFS, and be a major source of endogenous RS.



The mechanisms underlying replication-transcription conflicts are far more complex than the simple collision of two uncoordinated types of machinery. Ensuring replication/transcription co-directionality is fulfilled by the preferential initiation of replication at the transcriptional start site of highly transcribed genes, a mechanism exacerbated under conditions of RS<sup>87</sup>. Transcription may also be involved in conflicts with replication indirectly through the annealing of nascent RNA transcripts with their cognate DNA template, thereby forming structures known as R-loops<sup>88</sup>. Although present in homeostatic conditions throughout the genome, the mechanisms leading to R-loops toxicity in mammalian cells are still poorly understood. Widespread transcriptional awakening of simple repeats in *C. Elegans* leading to exaggerated R-loop formation and DNA damage has been reported<sup>89</sup>. However, R-loops toxicity was also demonstrated in human cells depleted from factors involved in replication fork progression<sup>90</sup>, DNA repair<sup>91</sup>, and chromatin remodeling<sup>92</sup>. How R-loops become toxic is still a matter of debate, with some evidence pointing to the stalling of replication forks due to secondary structures, such as hairpins, G-quadruplexes (G4)<sup>93</sup>, or H-DNA motifs<sup>94</sup>, formed by the displaced ssDNA at those regions. Alternatively, R-loops may form behind those forks under stress conditions and displace factors necessary for their processing<sup>91</sup>. Although R-loops have been linked to the formation of secondary DNA structures, the opposite has also been shown using synthetic G4 ligands stabilizing those structures, which led to an increase in R-loops levels and DNA damage<sup>93</sup>, thus highlighting the complex interplay and yet unknown regulatory mechanisms underlying this cross-talk.

## **2.3 Endogenous replicative stress, state of knowledge in DLBCL**

During B cell activation, the antibody diversification process is under the influence of the mutagenic activation-induced deaminase (AID) which transforms cytosine into uracil residues<sup>95</sup>, leading to its replacement by thymidine upon DNA replication. The exact mechanisms governing the tethering of AID to IgH and VDJ loci are still unknown, although their dependency on active transcription and ssDNA exposure through the formation of R-loops has been demonstrated<sup>96</sup>. In addition, AID has been shown to deaminate preferentially cytosines located in WRCY motifs<sup>97</sup>, thus explaining part of its physiological role at the IgH locus. In GC-derived B cell lymphomas, AID upregulation has been shown to trigger RS by causing off-target point mutations, encompassing the MYC loci, thus promoting its translocation event<sup>98</sup>. Additionally, centroblasts (the early activated form of B cells) are characterized by an extremely short cell cycling time of 6-7 hours, probably one of the shortest of the mammalian world<sup>99</sup>. The challenge of conducting a proper replication program in light of such fast cell cycling renders them prone to endogenous RS.

The proto-oncogene c-MYC located on Chr8q is translocated with an IGH enhancer in virtually all Burkitt lymphomas and 5-20% of DLBCL cases<sup>100</sup>. However, the MYC protein is detected as overexpressed in 30-40% of DLBCLs by IHC<sup>101</sup>, while it is nearly undetectable in normal germinal center B cells (<5% positive cells)<sup>102</sup>, indicating the presence of additional epigenetically-driven upregulation mechanisms in DLBCL cases lacking IGH/MYC translocations. MYC overexpression was shown to trigger RS through different mechanisms and to be associated with more aggressive lymphomas and solid tumors. It has been shown to trigger genome-wide bursts in transcription by promoting RNA pol II elongation through the recruitment of factors involved in its assembly<sup>103</sup>, thus increasing the likelihood of collision events between the replication and transcription machinery. Recent studies on B cells have also identified a subset of fragile sites occurring near replication origins, known as early replication fragile sites (ERFS), sensitive to MYC-induced replicative stress<sup>104</sup>. These ERFS could explain the occurrence of chromosomal breaks at BCL2, BACH2, and FOXP1 loci and their recurrent translocation in lymphoma. Apart from its role in promoting transcription-replication conflicts, MYC has been shown to facilitate the maturation of pre-RC complexes by inducing the decondensation of chromatin at its target loci, thereby promoting CDC45 and GIN complex recruitment and origin firing<sup>105</sup>. ATR and WEE1 inhibitors are currently being studied in MYC-driven and CDKN2A-deleted lymphomas which are particularly sensitive to RS<sup>106,107</sup>.

## 2.4 Heterochromatin regions as replicative stress hotspots

Genomic DNA is folded around octameric proteins known as histone complexes. Histone core subunits are composed of two H2A, H2B, H3, and H4 copies around which DNA is wrapped 2.3 times before folding itself again around the adjacent histone complex. These DNA: protein structures are known as nucleosomes, the fundamental subunits of the chromatin. Post-translational modifications of the tails of histone proteins, along with their interaction with other structural scaffold proteins, dictate the state of chromatin condensation. Chromatin compaction deserves, in addition to its structural role, a central function in restraining DNA accessibility to factors involved in gene expression and replication<sup>108</sup>. Therefore, genomic regions harboring genes actively transcribed are usually linked to a less condensed state of the chromatin, known as euchromatin. In opposition, the more compacted state of the chromatin, known as heterochromatin, is linked to repressed genes, repetitive DNA regions, and chromosomal extremities. Heterochromatin may also be functionally divided in 2 groups according to its propensity to be decondensed in different cell types and states, into facultative or constitutive heterochromatin. Facultative heterochromatin is usually associated with H3K27me3 deposits, which deposition is ensured by the methyltransferase activities of EZH1/2, the catalytic subunits of the Polycomb repressive complex 2 (PRC2), and is associated with gene expression plasticity, imprinting, and X chromosome inactivation<sup>109</sup>. In contrast, constitutive heterochromatin is enriched in H3K9me3 deposits and associated with lesser variability across cell types and states, and is usually considered to carry a more structural function<sup>110</sup>. Of course, this dichotomous vision eludes the many exceptions in which H3K9me3 changes influence cell identity and H3K27me3 changes may influence genome stability. Long envisioned as strongly associated with gene silencing, H3K9me3 deposits were shown to ensure gene expression stability when located at the 3' ends of transcribed genes in euchromatin regions<sup>111</sup>, and to tune the expression of some DNA repeats involved in chromosomal stability<sup>112</sup>.

Epigenetic modifiers are increasingly recognized as therapeutic targets, not only able to modulate critical gene expression, but also to elicit genotoxic effects in cancer cells. 5-azacytidine and its derivate 5-aza-2'-deoxycytidine are potent inhibitors of DNA methyltransferase 1 (DNMT1) leading to transcriptomic modifications<sup>113</sup>, cell differentiation<sup>114</sup>, oxidative stress<sup>115</sup>, and transcription-replication conflicts from global genome demethylation<sup>116</sup>. The latter being due to the widespread transcriptional awakening of genes and DNA repeats. The depletion of certain histone-modifying enzymes such as SETDB1

also demonstrated genotoxicity in mouse fibroblasts<sup>117</sup> and epithelial cells<sup>118</sup>. Mostly reported as triggering cell-intrinsic inflammatory responses through the upregulation of transposable elements (TEs) and the activation of dsRNA sensing pathways, the potential genotoxic effect of SETDB1 downregulation, along with the other five H3K9me3 methyltransferases, has not been studied in cancer cells so far. Recently, the loss of the PRC1 complex has been shown to enhance chromosomal instability (a form of genome instability affecting chromosome structure, segregation, and ploidy) in uveal melanoma. PRC1 is recruited to H3K27me3 loci to ubiquitylate H2A and K119 histones and lock the repression of those sites. Therefore, epigenetic modifications at facultative heterochromatin regions may trigger genomic instability and promote cancer progression. The involvement of transcription factors in genome stability, beyond the ones regulating the expression of DNA repair genes, has rarely been documented. One exception is the depletion of the transcriptional co-activator BCL3, which was recently reported to trigger DNA damage in adult T cell leukemia (ATL)<sup>119</sup> and colorectal cancer (CRC)<sup>120</sup> through yet unsolved mechanisms involving a reduction in homologous recombination DNA repair.

## **2.5 Replicative stress, a source of cell-intrinsic inflammatory responses**

The first evidence of a link between genome instability and inflammation came from the study of Ataxia Telangiectasia (AT) patient samples in which ATM deficiency led to cytosolic DNA accumulation and type I interferon (IFN) responses<sup>121</sup>. This pioneering work also demonstrated that irradiation-induced DNA damage was able to trigger the same type of inflammatory response. Genomic instability has been shown to induce cell-intrinsic inflammatory responses through the shedding of DNA fragments into the cytosol<sup>122,123,124</sup> or the formation of micronuclei<sup>125,126</sup>. Micronuclei are the consequence of chromosomal missegregation events occurring during mitosis, in which lagging chromosomes are trapped in these extranuclear structures. Their rupture leads to the release of dsDNA in the cytosol which is recognized by an enzyme with cyclic GMP-AMP synthase activity known as cGAS<sup>127</sup>. The product of this enzymatic reaction, cGAMP, is then recognized by a transmembrane protein associated with the endoplasmic reticulum called STING, leading to its relocation in the Golgi apparatus. This event triggers a signaling cascade culminating in the phosphorylation of IRF3 and the release of NF- $\kappa$ B from the IKK complex through the recruitment of the TANK-binding kinase 1 (TBK1) and I $\kappa$ B kinase, respectively<sup>128</sup>. Phospho-IRF3 and NF- $\kappa$ B then translocate into the nucleus to their target genes, encompassing IFN $\alpha$  and IFN $\beta$ . In solid tumors, micronuclei-related IFN responses are now recognized as an important mechanism in the metastatic process of these tumors as they appeared to be coupled with the activation of epithelial-to-mesenchymal transition (EMT), as if inflammation was a price to pay in the route of cancer progression, although the mechanisms of this coupling are still unknown<sup>125</sup>.

The direct shedding of dsDNA fragments from the nucleus into the cytosol was described in Aicardi-Goutières syndrome (AGS)<sup>129</sup>, an autosomal recessive hereditary disease first described in 1984 by two French neuropediatricians at Necker's hospital<sup>130</sup>. Children affected by this neuro-inflammatory disease presented lymphocytosis and increased levels of IFN $\alpha$  in their cerebrospinal fluids in the absence of any ongoing infection. These patients were shown to carry germline mutations in genes encoding DNA or RNA nucleases, such as TREX1, SAMHD1, and RNase H2 genes<sup>131</sup>. SAMHD1 was found to enhance the exonuclease activity of MRE11<sup>129</sup>, a protein responsible for the degradation of nascent DNA during RS, thereby facilitating replication fork restart<sup>129</sup>. Therefore, impaired SAMHD1 activity leads to aberrant processing of replication forks and decreased MRE11 exonuclease activity leading to non-cleaved newly synthesized ssDNA fragments. However, MRE11 also carries an endonuclease

activity that is not impaired by SAMHD1 deficiency, leading to the accumulation of ssDNA fragments. Nuclear export mechanisms underlying ssDNA/dsDNA export are still poorly understood. TREX, an exonuclease located at the outer surface of the nuclear membrane, is also mutated in some cases of AGS leading to its decreased activity and ssDNA cytosolic accumulation<sup>132</sup>. The cancer relevance TREX was shown in irradiated breast and colorectal cells, in which it dictated the level of cytosolic DNA and influenced abscopal anti-tumor responses in mice models<sup>133,134</sup>. SAMHD1 and TREX alterations have been reported in various tumor types<sup>135,136</sup>. MUS81, an endonuclease involved in DNA cleavage at stalled replication forks, has recently been shown to be a source of cytosolic dsDNA at the early stages of prostate cancer development<sup>137</sup>. These examples illustrate the duality of genomic instability as promoting cancer development on one hand and triggering inflammatory responses on the other, thus opening new opportunities in the field of cancer therapeutics.

### **3. KRAB zinc finger proteins as maintainers of constitutive heterochromatin memory**

#### **3.1 Transcription factors in cancer, an overview**

Cancer is considered a disease of cell uncontrolled proliferation and invasion, led by the key concept of clonal evolution, in which genetically altered cells outweigh their basic function to become autonomous and freed from outgrowth limitations in a multi-step process. TFs are a group of approximately 1'200 proteins in the human genome devoted to the regulation of gene expression<sup>138</sup>, each contributing to the architecture of transcriptional networks at various extents but also to higher-order chromatin structures, ultimately defining the cell identity and function. Although extremely complex, decades of scientific research shed the light on the apparent hierarchical organization and context-dependency of these proteins, with the alteration of some being sufficient to trigger neoplastic transformation (known as “cancer drivers”), serve tumor suppressive functions, or contribute to particular cancer hallmarks. Most cancer-driving gene alterations affecting TFs have been reported in hematological malignancies, which is not surprising owing to the phenotypic plasticity of hematopoiesis aiming at replenishing throughout life blood cells with highly diverse functions, which highlight the critical role of epigenetic regulation in blood cell differentiation. These alterations lead to the dysregulation of normal differentiation programs and to increased proliferative and metabolic capacities in precancerous hematopoiesis<sup>139</sup>, but also enhanced genomic instability as they may impair the expression of DNA repair genes, thus fostering the acquisition of additional mutations and cancer progression. Furthermore, these types of driving initial events have also been found to occur in some solid tumors where primary alterations of signaling kinases and DNA repair genes are more frequently at the forefront of cancer-initiating events. TFs shape gene expression throughout cancer evolution and contribute to numerous cancer hallmarks by controlling networks involved in processes such as epithelial-to-mesenchymal (EMT) transition, immune interactions, replicative immortality, metabolism, and cell death. In addition to their inherent combinatorial complexity, some TFs may also become essential in transformed cells, while devoid of critical function in their healthy counterparts, a phenomenon designated as “non-oncogene addiction”<sup>140</sup>.

Beyond genetic alteration, some TFs may contribute to cancer progression without being modified. This is exemplified by some critical TFs being either upregulated in specific

neoplastic contexts, such as Myc or BCL6 in the case of DLBCL. Even more challenging is the duality that certain TFs may harbor, as some serve tumor suppressive functions in non-transformed cells, and opposite functions in cancer. Perhaps one of the best illustrations of this duality is TP53, the so-called “guardian of the genome”, and the most recurrently mutated gene in cancer interesting approximately 50% of all tumors<sup>141</sup>. This TF is the master regulator of a network involved in safeguarding genome integrity, through the upregulation of the DNA repair machinery and cell-cycle regulators, which becomes active upon DNA insults. Alteration of the TP53 gene is one of the initial events of numerous carcinogenic processes, ultimately leading to the accumulation of additional DNA lesions at critical genomic loci. Nevertheless, TP53 mutations not only diminish its normal tumor suppressive function but also confers newly acquired oncogenic functions, in addition to its association with a dominant negative phenotype, in which altered TP53 impairs the activity of its wild-type allelic partner<sup>142</sup>. Beyond gene expression regulation, increasing work is showing that some TFs and their interacting epigenetic modifiers are involved in the maintenance of genomic integrity, making them interesting therapeutic targets as DNA damage sensitizers.



### **3.2 KZFPs and their TE targets, a source of epigenetic diversity and cancer plasticity**

KRAB zinc finger proteins are the largest family of transcription factors in the human genome, counting nearly 400 members<sup>138</sup>, half of which are located on chromosome 19<sup>143</sup>. They are named according to their defining domain: the Krüppel-associated box (KRAB), which is followed by an array of DNA-binding C2H2 zinc fingers. Located at the C-terminus of KZFPs, these arrays are composed of multiple repeats of C2H2 zinc fingers (on average 3-41) providing them a large repertoire of DNA target sequences<sup>144</sup>. The KRAB domain is a conserved sequence of 75 amino acids found at the N-terminus of these proteins<sup>145</sup>. It is responsible for the transcriptional repression activity of KZFPs, as it was shown to recruit the scaffold protein KAP1 (also known as TRIM28) which recruits a enzymatic machinery comprising nucleosome remodeling, histone deacetylase, and histone and DNA methyltransferase activities leading to heterochromatin deposition at their target loci<sup>146,147</sup>. Conditional KAP1 KO experiments in mice B cells revealed it to be required for late B cell maturation, by regulating genes involved in cell migration, differentiation and apoptosis<sup>148</sup>. Paradoxically, specific KZFPs have been shown to activate gene expression in a KAP1-dependent manner through suspected locus-specific interactions, such as the activation of adipogenic genes by ZFP30<sup>149</sup>. A fraction of KZFPs also carries a variant KRAB domain, lacking the essential amino acid residues for KAP1 recruitment, most of which are still not characterized<sup>150</sup>, with the notable exception of variant KRAB-containing ZNF398 also known to carry an additional domain of unknown function (DUF), which was showed to activate the expression of pluripotency genes in human embryonic stem cells, through the recruitment of the histone acetyltransferase EP300<sup>151</sup>.

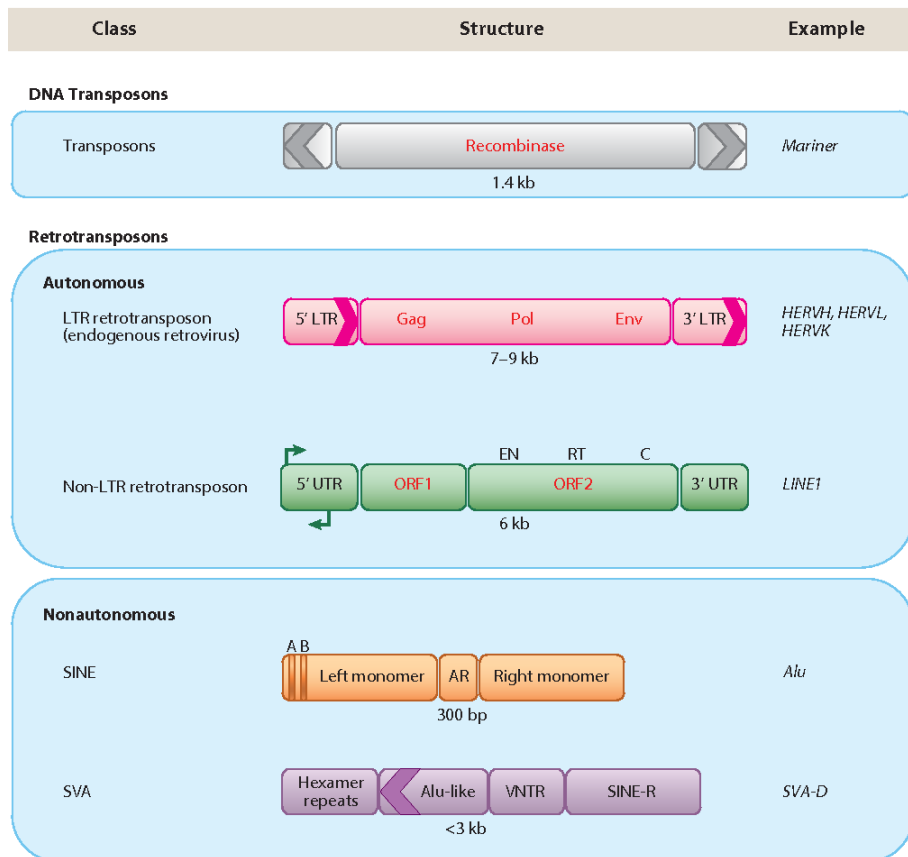
KRAB zinc finger proteins are involved in many biological processes, including embryonic development, cell differentiation, and immune responses. One of their most characterized activities was the repression of mobile genetic elements, known as transposable elements (TEs), during early embryogenesis<sup>152,153,154</sup>, thus KZFPs are considered to be a protective mechanism against the mutagenic potential of these elements. However, only a few hundred of TE copies are still able to retrotranspose<sup>155</sup>, contrasting with the hundreds of thousands of KZFP-binding sites on those elements<sup>156</sup>. It was shown that KZFPs not only repress the transcription of TEs, but also participate to regulate the cis-regulatory activity of some TEs during tissue differentiation<sup>153</sup> or germ cells<sup>157</sup>. It is however broadly accepted that once their H3K9me3 and DNA methylation have been deposited, they are not necessary to maintain this state of repression in adult tissues<sup>158</sup>, with only a few exceptions showing the opposite<sup>159</sup>. If

this is also the case in cancer cells in which DNA methylation and chromatin alterations are frequent is not known. Some studies have suggested that KZFPs display tumor-suppressive activities for their vast majority by repressing the expression of genes that promote cell proliferation, survival, and EMT<sup>160</sup>. However, a smaller set of studies have shown the opposite for some KZFPs promoting cell proliferation<sup>161</sup>, survival<sup>162,163</sup>, epithelial to mesenchymal transition (EMT)<sup>164</sup>, chemoresistance<sup>165</sup>, and metastasis<sup>166,160</sup>, and have been proposed to promote disease in lymphocytic leukemia<sup>167</sup>, lung<sup>168, 169</sup>, colorectal<sup>170</sup>, and ovarian cancer<sup>171</sup>. Most of the cancer studies published so far focused on conserved mammalian or evolutionary older KZFPs, occulting the potential role of primate- and hominoid-specific in carcinogenesis which account for half of KZFP genes, as most are the product of evolutionary recent gene or segment duplications. APAK (also known as ZNF420) is a mammalian KZFP known to downregulate the activity of TP53 through a direct protein-protein interaction<sup>172</sup>. This example illustrates that KZFPs may harbor functions beyond their sole transcriptional regulatory function and that, although evolutionary recent, some may regulate essential aspects of cancer progression. This TP53 post-translational regulatory function involving KZFPs is not unique, as ZER6 was shown to promote TP53 ubiquitination similarly to KAP1<sup>173</sup>. Overall, these studies suggest that the role of KZFP in cancer is context-dependent and can vary for different KZFPs. Interestingly, two canonical primate-restricted KZFPs paralogs, ZNF587 and ZNF417, were recently shown to prevent deleterious IFN responses in brain organoids, in addition to regulating the expression of neuron-specific developmental genes<sup>154</sup>. Owing to the anti-tumor effect of acute IFN responses, the therapeutic potential of triggering such responses through KZFP downregulation in cancer cells should deserve further emphasis.

Retrotransposons are repetitive DNA sequences that can jump around the genome using a “copy and paste” mechanism involving the retrotransposition of an RNA intermediate. Retrotransposons compose nearly half of the genome and are the product of evolutionary waves of retrotransposition, with, however, a few hundred elements still able to spread during embryonic genome activation or gametogenesis<sup>155</sup>, thus shaping the epigenome of newborns at a frequency of one new insertion every 1/20 births for Alu, 1/200 for LINE1, 1/900 for SVA retroelements, respectively<sup>174</sup>. TEs contribute to establish and expand gene regulatory networks, as nearly 20% of all TF binding sites are embedded within TEs<sup>175</sup>. Some TE copies have also been selected as a source of alternative gene promoters, splicing variants, or alternative 3'ends, thus contributing to cell plasticity.

Retrotransposons are divided into 4 families based on sequence structure similarities: the long terminal repeat (LTR) elements (also known as endogenous retroviruses, ERV), the long-

interspersed elements (also known as LINE1), the short-interspersed nuclear elements (SINE), and the SINE-VNTR-Alu-like (SVA) elements. LTR/ERV retrotransposons account for approximately 100'000 copies making up 8% of the human genome. They are remnants of ancient retroviral infections that have become integrated into the genome of placental mammals and spread across successive waves coinciding with mammalian up to hominoid radiation events<sup>176</sup>. Non-LTR retrotransposons do not contain LTRs and instead use a poly-A tail for integration<sup>177,178</sup>. LINE1 account also for approximately 500'000 copies making up 17% of the human genome, being, therefore, the most prevalent<sup>179</sup>. SINEs, which are composed of the human-specific Alu and mammalian-specific MIR elements are shorter than LINEs, and rely on the reverse transcription machinery of the few hundred active LINE1 for their mobility. SINEs account for more than 1 million copies throughout the genome making up 11% of its DNA content. Alu elements are enriched in TP53 binding sites, with more than 400'000 potential sites, thus making the spread of Alu elements a major contribution to TP53 regulatory network evolution<sup>180</sup>. The hominoid-specific SVA family is composed of surrounding 3'000-5'000 integrants and accounts for nearly 0.3% of the genome, and similarly to SINEs, hijacks the LINE1 machinery for its own spread. A small fraction of the human genome (3%)<sup>181</sup> is also composed of old transposons unable to move in the human genome (non autonomous). Nevertheless, they are still very active (autonomous) in other living organisms, in which they transpose through a cut and paste mechanisms and expand through homology-dependent DNA repair of their initial insertion site. At the population level, TEs are a source of genetic diversity, allowing the emergence of new traits indispensable for adaptation to changing environment or other types of selective pressures.



**Figure 4 | Structure of transposable elements.** A consensus example of each main TE class is depicted with their respective average length indicated below<sup>182</sup>.

Their mutagenic potential can disrupt genes and regulatory regions leading to new protein products and novel cis-regulatory sequences, although most of their insertions are silent or lead to small variations in gene regulation. Beyond insertional mutagenesis, TEs can disrupt genome integrity through homologous allelic and non-allelic homologous recombination due to their sequence similarity or the potential production of genotoxic R-loops, thus stressing the role of chromatin compaction to limit it<sup>183,184</sup>. LINE-1 retrotransposon activity has been shown across many cancer types, including breast, ovarian, and colorectal cancer, and has been associated with poor prognosis and resistance to chemotherapy, by promoting chromosomal instability through insertional mutagenesis<sup>185</sup> and promoted driver gene rearrangements<sup>186</sup>, or through the genotoxic activity of the LINE1-encoded endonuclease ORF2 protein<sup>187</sup>.

Other studies have suggested that TEs can also carry a tumor-suppressive role through the activation of innate immune sensing subsequently enhancing the recognition of cancer cells by the adaptive immune system, such as upon the depletion of the methyltransferase SETDB1 in melanoma and lung cancer mice models<sup>188</sup>. In the latter, the transcriptional awakening of TEs, more specifically ERV elements, recognized by the RIGI-MAVS dsRNA sensing pathway, the

counterpart of the cGAS-STING pathway for dsRNA sensing, enhanced the efficacy of immune checkpoint inhibitors, through the production of an IFN response and the cell surface exposure of HLA-I loaded TE-derived peptides.

Overall, the role of TEs in cancer is complex, as they can act as drivers of tumor progression by promoting genomic instability, enhancing the expression of oncogenes through their cis-regulatory potential, or enhancing immune surveillance, thus promoting immune anti-tumor responses. The cis-regulatory potential of TEs in driving oncogene expression has been demonstrated in acute myelogenous leukemia<sup>189</sup>, where elements of the ERV family have been shown to drive the expression of genes necessary for leukemic cell metabolism and proliferation. Most interestingly, the sequence similarity of these cis-regulatory elements allowed their broad downregulation using CRISPR interference with sgRNAs able to target a hundred sites across the genome simultaneously, thus offering new therapeutic opportunities using gene silencing to target oncogenic gene regulatory networks.

## Aims of the thesis

The role of KZFPs, the largest family of transcription factors in the human genome, in the pathogenesis of DLBCL has not been addressed so far. As nearly half of them are primate-restricted and don't present recurrent mutations in cancer, they are often overlooked in this pathologic context. Nevertheless, their TE targets are increasingly being recognized as important actors of tumorigenesis, and therefore some of their KZFP controllers are also presumably involved in this process. I decided to start by linking their expression level across DLBCL samples with patient outcomes, in a uniformly treated cohort, in order to identify and characterize pro-oncogenic and tumor-suppressive gene candidates, for further *in vitro* studies. The most important objective of this study was to unveil the possible contribution of KZFPs to DLBCL phenotype.

## Contribution

The following manuscript is currently under peer review. It constitutes the main body of my thesis. I designed the study, obtained data accesses, and performed alone most part of the experiments with the exception of CUT&Tag, TrAEL-Seq, and immunopeptidomic experiments which were performed with the help of Dr. O. Rosspopoff, Dr. R. Forey, and Prof. Michal Bassani's team, respectively. I benefited from the technical support of S. Offner. Bioinformatic analyses were performed by J. Carlevaro-Fita, E. Planet regarding pre-processing of RNA-Seq, CUT&Tag and TrAEL-Seq data. HLA predictions and raw data analyses of MS sequencing were performed by Florian Huber from the Laboratory of Prof. Michal Bassani. I interpreted the data and wrote the manuscript with corrections from all the authors.

# **KRAB zinc finger proteins ZNF587/ZNF417 protect lymphoma cells from replicative stress- induced inflammation**

Filipe Martins<sup>1,2,3</sup>, Olga Rosspopoff<sup>1</sup>, Joana Carlevaro-Fita<sup>1</sup>, Romain Forey<sup>1</sup>, Sandra Offner<sup>1</sup>, Evarist Planet<sup>1</sup>, Cyril Pulver<sup>1</sup>, HuiSong Pak<sup>3,4,5</sup>, Florian Huber<sup>3,4,5</sup>, Justine Michaux<sup>3,4,5</sup>, Michal Bassani-Sternberg<sup>3,4,5</sup>, Priscilla Turelli<sup>1</sup> and Didier Trono<sup>1\*</sup>

<sup>1</sup> School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

<sup>2</sup> Service and Central Laboratory of Hematology, and <sup>3</sup>Department of Oncology, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.

<sup>4</sup> Agora Cancer Research Centre, Lausanne, Switzerland.

<sup>5</sup> Ludwig Institute for Cancer Research, University of Lausanne, Lausanne, Switzerland.

\*: Corresponding author and lead contact

didier.trono@epfl.ch



# INTRODUCTION

Diffuse large B cell lymphoma (DLBCL), the most common type of lymphoid malignancy<sup>190</sup>, is plagued by relapse or refractoriness to first-line therapy in nearly fifty percent of cases<sup>37</sup>. DLBCL arises from the neoplastic transformation of mature B cells during the final steps of their differentiation into antibody-secreting plasma cells or memory B cells<sup>191</sup>. This program is normally under tight control by epigenetic mechanisms ensuring the coordinated activation of specific gene expression networks<sup>192</sup>, but it can be dysregulated by aging<sup>193</sup>, mutations in critical genes<sup>194</sup> and other environmental influences<sup>195</sup>, leading to lymphomagenesis. Three major molecular subtypes of DLBCL with prognostic relevance<sup>196,10</sup> have been delineated based on similarities between the transcriptome of these tumors and their putative cell of origin (COO). Germinal center B cell (GCB) and activated B cell (ABC) DLBCLs are thought to derive from B cells blocked at the centroblast and plasmablast stages, respectively, while the transcriptome of so-called “unclassifiable” DLBCL does not evoke any of these two differentiation stages.

Genomic instability is a primary driver of the clonal evolution and diversification of cancer cells<sup>125,197</sup>. It frequently results from errors occurring during genome replication, which starts with the assembly of pre-replicative complexes (pre-RCs) during the G1 phase of the cell cycle, an event known as origin licensing<sup>52</sup>. During early S phase, pre-RCs recruit the DNA polymerase and elongation factors to form the replication fork machinery. Subsets of licensed replication origins then become activated in a sequential and spatially regulated manner<sup>198</sup>, allowing the bidirectional progression of replication forks across the genome. Conducting a proper replication program in the presence of short cell cycling<sup>61,62</sup>, intercurrent mutagenic events<sup>63</sup> and exacerbated oncogene-induced transcriptional activity<sup>64</sup> is challenging, and accordingly DLBCL cells are prone to replicative stress (RS), characterized by the stalling and potential collapse of replication forks leading to genomic instability<sup>65</sup>.

Increasing evidence suggests a central role for chromatin compaction in the selection and timing of activation of DNA replication origins<sup>108</sup>. Heterochromatin regions are known to be late-replicating, a phenomenon attributed to their high content in repetitive DNA sequences, a major source of RS as they provide hotspots for homologous recombination, secondary DNA structures, and the formation of RNA:DNA hybrids<sup>89,93,199,183,184</sup>. Pericentromeric regions and telomeres, which are enriched for satellite and simple repeats, are the most heterochromatin-dense regions of the genome. Nevertheless, some 4.5 million transposable elements (TEs) sprinkled over the rest of the human genome also contribute significantly to its heterochromatin

landscape. Most human TEs are retroelements belonging to the LTR (long terminal repeat)-containing endogenous retroviruses (ERV), long and short interspersed nuclear elements (LINE and SINE, respectively) or SVA (SINE-VNTR-Alu-like) families, which replicate by a “copy-and-paste” mechanism through reverse transcription of an RNA intermediate and integration of its DNA copy elsewhere in the genome. TEs are motors of genome evolution<sup>179</sup>, and a major source of *cis*-acting regulatory sequences that shape gene regulatory networks that play in broad aspects of human biology<sup>159,200,181</sup>. The disruptive potential of TEs has led to the evolutionary selection of many protective mechanisms, whether to prevent transposition during genome reprogramming in the germline and early embryogenesis or to control the transcriptional impact of TE-embedded regulatory sequences. Repression of TE loci via DNA methylation and heterochromatin formation through methylation of lysine 9 of histone 3 (H3K9me3), stand prominently amongst these mechanisms, and Krüppel-associated box (KRAB) zinc finger proteins (KZFPs) amongst their mediators<sup>158</sup>.

KZFPs constitute the largest family of transcription factors (TFs) encoded by higher vertebrates, with nearly 400 members in human alone, more than a third of them primate-restricted and the products of recent gene duplication events<sup>156,138</sup>. KZFPs bind DNA in a sequence-specific manner through their C-terminal arrays of zinc fingers, and a large majority of human KZFPs have TE-embedded sequences as primary genomic targets<sup>156</sup>. Most act as transcriptional repressors via the KRAB domain-mediated recruitment of KAP1 (KRAB-associated protein 1, also known as tripartite motif protein 28 or TRIM28)<sup>150,201,146</sup>, which serves as a scaffold for the assembly of a heterochromatin-inducing complex comprising nucleosome remodeling, histone deacetylase, and histone and DNA methyltransferase activities<sup>150</sup>. Rather than just another silencing mechanism aimed at preventing TE mobilization, KZFPs are increasingly recognized as facilitating the domestication of TEs regulatory potential<sup>202,156</sup>. Indeed, KZFPs not only repress TEs immediately after genomic activation but also corral their transcriptional influences at various stages of development and in adult tissues<sup>149,157,203,204,205</sup>. Thus, KZFPs both minimize the genotoxic potential of TEs and partner up with their genomic targets to shape lineage- and species-specific transcription regulatory networks<sup>153,152</sup>.

KZFPs have been proposed to exert collectively tumor suppressive functions across many tumor types<sup>160</sup>. However, individual family members have been linked to carcinogenesis through the deregulation of genes implicated in cell cycle progression<sup>161</sup>, apoptosis<sup>162,163</sup>, epithelial to mesenchymal transition (EMT)<sup>164</sup>, chemoresistance<sup>165</sup>, and metastasis<sup>166,160</sup>, and have been proposed to promote disease in lymphocytic leukemia<sup>167</sup>, lung<sup>168, 169</sup>, colorectal<sup>170</sup>,

and ovarian cancer<sup>171</sup>. In addition, their main cofactor KAP1 functions as an E3-ubiquitin ligase capable of inactivating p53<sup>206</sup>, and the KAP1-recruited SETDB1 methyltransferase was recently found to foster immune evasion in melanoma and lung cancer by repressing TE transcription, a source of dsDNA/dsRNA species sensed by intracellular innate immunity pathways and of endogenous viral peptides with immunogenic properties<sup>188,207</sup>.

Here, we took advantage of a large dataset of DLBCL patients with extensive clinical and molecular documentation<sup>208</sup> to explore the potential role of KZFPs in this hematological malignancy. This led us to discover that ZNF587 and ZNF417, a pair of primate-specific KZFP paralogs targeting evolutionarily recent TEs and so far mostly known for their involvement in human brain development and the control of neuronal gene expression, protect DLBCL cells from excessive replicative stress associated with genome-wide alterations of the heterochromatin landscape, thus safeguarding these cells from cell-intrinsic inflammatory responses.

# RESULTS

## A DLBCL transcriptome-wide association study reveals KZFPs as predictors of poor prognosis

To probe the potential involvement of KZFPs in DLBCL pathogenesis, we looked for a correlation between their levels of expression and disease outcome. For this, we used published RNA-seq data from 633 fresh-frozen paraffin-embedded DLBCL samples collected before the initiation of rituximab-based therapy<sup>208</sup>. We conducted a univariate Cox proportional hazards regression analysis on the transcriptome of these samples to identify genes, the expression of which was either positively or negatively correlated with survival. Out of 23'560 expressed genes (see methods), 1'530 (6.5%) were found to be significantly associated with either a better (n=950, 62.1%) or a worse (n=580, 37.9%) outcome (Fig. 1a). KZFP genes were significantly enriched in prognosis-associated genes with 41 out of 395 (10.4%) expressed KZFPs associated with survival (1'530/23'560 vs. 41/395; *Fisher's exact test*,  $p = 0.0038$ ), 39 of them negatively. This stood in sharp contrast with the observed distribution of hazards ratios (HRs) for all other TFs or remaining protein-coding genes, where good-prognosis hits dominated (Fig. 1b). Corroborating our results, a recent analysis of the same cohort found 12 of these 39 KZFP genes to be part of a KZFP-enriched poor-prognosis gene expression signature<sup>209</sup>.

Among the 41 *KZFP* gene hits of our screen, 39 were protein-coding, 19 of them primate-specific and expressed at higher levels than their evolutionary older counterparts in all disease subtypes (Fig. 1c)<sup>156</sup>, suggesting a possible (COO-independent) pro-oncogenic role for these proteins despite their young evolutionary age. *ZNF662* and *ZNF354C*, the two better-prognosis *KZFP* genes identified by our analysis, reside on chromosomes 3 and 5, respectively. Interestingly, the long arm of chromosome 19 (Chr19q) harbors 32 of the 39 (82.1%) poor prognosis *KZFP* genes, a significant overrepresentation of this genomic region as the entire chromosome 19 only encodes slightly more than half of all KZFPs (231/395, 58.5%; *Fisher's exact test*,  $p = 0.0034$ )<sup>210,1211</sup>. To evaluate if these *KZFP* genes were part of a single gene expression module, we asked which were systematically co-expressed in DLBCL samples. This led us to identify a subset of 18 *KZFPs* displaying a strong co-expression pattern. Fifteen of these were located within the C11 *KZFP* gene cluster located on the sub-telomeric region of chromosome 19 (chr19q13.43) and the other 3 within the nearby C9 cluster (Fig. 1d)<sup>152</sup>. This

prompted us to characterize the DLBCL subset overexpressing these coregulated unfavorable *KZFP* genes.

## **A *KZFP* gene cluster defines a DLBCL subset with cell-autonomous growth features**

Through unsupervised hierarchical clustering of DLBCL samples based on the expression of these 18 co-expressed *KZFP* genes, we could delineate a subset encompassing about a quarter of all DLBCL samples ( $n=145$ , 23%, Fig. 2a), to which we will refer to hereafter as *KZFP*<sup>High</sup> DLBCLs. To characterize differences between these samples and their *KZFP*<sup>Low</sup> counterparts, we conducted a gene set enrichment analysis (GSEA)<sup>212,213</sup> for pathways catalogued in the Hallmark gene sets<sup>214</sup>. *KZFP*<sup>High</sup> samples were characterized by a proliferative and highly metabolic state, as suggested by the upregulation of genes involved in cell cycle regulation, DNA repair, mitotic spindle assembly, oxidative phosphorylation and Myc response. Conversely, *KZFP*<sup>Low</sup> samples were characterized by the activation of interferon (IFN)/inflammatory response and EMT (epithelial-to-mesenchymal transition) gene sets (Fig. 2b).

Using the metadata available for this patient cohort<sup>208</sup>, we determined that individuals presenting with *KZFP*<sup>High</sup> DLBCL had survival rates twice shorter than those with a *KZFP*<sup>Low</sup> profile (Fig. 2c). *KZFP*<sup>High</sup> patients were slightly older ( $65 \pm 13.2$  vs  $59 \pm 16.2$  years) and presented more frequently DLBCLs of the ABC subtype (47.2% vs. 34.7%) (Fig. 2d), but were diagnosed at similar disease stages (Fig. 2d). To assess the independence of these *KZFP* risk groups and the outcome predicted by age, stage at diagnosis and COO, we conducted a multivariate Cox regression analysis, which confirmed that *KZFP* risk groups remained significantly associated with survival ( $HR= 1.31$ ,  $p = 0.04$ ). Intriguingly, *KZFP*<sup>High</sup> DLBCLs did not appear to present a lower response rate to first-line rituximab-based regimens (Fig. 2d), suggesting that patients in this group were instead more inclined to relapse after first achieving remission. Although the data was not available to confirm directly this point, a Kaplan Meier analysis revealed that the steepest decrease in the survival rate of *KZFP*<sup>High</sup> patients occurred during the first 2.5 years following diagnosis (Fig. 2c), which is the most at-risk period for DLBCL relapse<sup>215</sup>. Furthermore, analysis of the HR by stratifying *KZFP*<sup>High/Low</sup> groups showed that patients in the *KZFP*<sup>High</sup> group had a higher hazard of death peaking in the first 6 months

after diagnosis, which remained higher than their KZFP<sup>Low</sup> counterparts up to 2-3 years, supporting the suspicion of higher incidence of relapses in the former group (Fig. 8a).

This led us to hypothesize that *KZFP* expression may contribute to disease relapse not by causing primary chemo-resistance of the bulk population of lymphoma cells, but rather by contributing to their clonal diversity, thus facilitating the emergence of treatment-resistant subclones, and by fostering the immune evasion of residual disease. To test this hypothesis, we applied a bulk RNA-seq deconvolution method<sup>216</sup> to DLBCL samples as a proxy for comparing the tumor microenvironment (TME) cell compositions of KZFP<sup>High</sup> and KZFP<sup>Low</sup> subtypes. This revealed a significant depletion of both innate and adaptative immune cells in KZFP<sup>High</sup> compared with KZFP<sup>Low</sup> DLBCL samples (Fig. 2e). This drop in immune gene signatures concerned T cell subtypes, activated Natural Killer (NK) cells, and monocyte-derived cells, with the exception of activated dendritic cells and eosinophils. Thus, KZFP<sup>High</sup> DLBCLs triggered weaker immune responses than their KZFP<sup>Low</sup> counterparts. These apparent differences in KZFP<sup>High/Low</sup> TME were further supported by the analysis of the mutational pattern of the KZFP-defined DLBCL groups, with KZFP<sup>High</sup> samples displaying significant enrichment for a recently defined high-risk molecular gene signature linked to *MYD88* and *CD79B* mutations, together with an increased prevalence of *BCL2* and *CDKN2A* alterations (Fig. 8b)<sup>217,218</sup>. This mutational signature, known as the MCD subtype, has been shown to promote DLBCLs immune evasion by reducing antigen presentation and NK cell activation amongst other mechanisms<sup>219</sup>. In addition, the analysis of whole-exome sequencing (WES) data issued from the same samples indicated a higher number of copy number alterations (CNAs) amongst KZFP<sup>High</sup> DLBCLs, compatible with greater clonal diversity (Fig. 2f). Taken together, these results indicate that KZFP<sup>High</sup> DLBCLs are characterized by a proliferative, genetically unstable and cell-autonomous phenotype, as well as an increased frequency of genetic changes associated with aggressive forms of the disease.

## **ZNF587/417 depletion impairs lymphoma cell growth and viability**

The most highly expressed poor prognosis-associated *KZFPs* across DLBCL samples, *ZNF586*, *ZNF587B*, and *ZNF587*, are located in a 200kb region of the Chr19q13.43 cytoband in the vicinity of other poor-prognosis *KZFPs* (*ZNF417*, *ZNF814*, *ZNF776*, *ZNF671*, and *ZNF552*). As a first step to investigate the impact of these genes on DLBCL homeostasis, we performed proliferation assays in U2932 and OCI-Ly7 cells, respectively representative of ABC and GCB DLBCL, transduced with lentiviral vectors expressing small hairpin RNAs (shRNAs) against these 3 *KZFPs* and their respective paralogs, or a scrambled sequence as a negative control (Fig. 9a). *ZNF587B/ZNF814* and *ZNF417/ZNF587*, two pairs of paralogs with respectively 70% and 98% of sequence identity, could be simultaneously repressed with a single shRNA targeting shared homologous regions in both paralog pairs. Although highly expressed in these cells (Fig. 9b), *ZNF587B/814* could be knocked down without affecting the growth of the U2932 ABC or OCI-Ly7 GCB cell lines (Fig. 3a). In contrast, *ZNF586* depletion stopped the proliferation of U2932 but was less efficient in decreasing OCI-Ly7 cell growth, whereas *ZNF587/417* knockdown (KD) triggered abrupt growth arrest in both settings (Fig. 3a). This result was confirmed with a second shRNA targeting *ZNF587/417* with similar KD efficiency (Fig. 9a/c) and in two additional cells lines, with growth arrest induced in both HBL1 (ABC) and SUDHL4 (GCB) cells, for the latter with a marked reduction in cell viability (Fig. 3b). Most importantly, knockdown of *ZNF587* and *ZNF417* did not preclude the proliferation of normal diploid primary foreskin fibroblasts (Fig. 9d), as similarly observed in human embryonic stem cells<sup>154</sup>, suggesting that dependence on these two *KZFPs* is somewhat restricted to cancer cells.

In order to understand the mechanisms underlying this observation, we performed RNA sequencing (RNA-seq) on U2932 cells harvested 3 days after *ZNF587/417* KD. We identified 332 differentially expressed genes (DEGs, fold change > 2 and FDR <0.05), 178 (53.6%) of which were downregulated (Fig. 3c). Gene set enrichment analysis (GSEA) revealed UV response UP as the most affected hallmark amongst DEGs. It encompassed several cyclins (*CCND3*, *CCNE1*), cyclin-dependent kinase inhibitors (*CDKN1A*, *CDKN1B*, *CDKN1C*), and TP53-responsive genes (*TP53I3*) (Fig. 9e). Reminiscent of the form of cell death induced by irradiation<sup>220</sup>, Annexin V and Propidium Iodide (PI) staining revealed an increase in PI+/AV- cells, a pattern indicative of necrosis, upon *ZNF587/417* KD in U2932 and most prominently SUDHL4 cells (Fig. 3f). This form of cell death appeared to be programmed, as the phosphorylation of MLKL, the downstream effector of the canonical necroptotic pathway, was

detected by western blot (data not shown). Thus, ZNF587/417 depletion impaired proliferation and triggered necroptosis death across DLBCL cell lines.

### **ZNF587/417 depletion leads to a broad downregulation of KZFPs and alters the heterochromatin landscape of lymphoma cells**

ZNF587 and ZNF417 are both KAP1-binding KZFPs that secondarily recruit the methyltransferase SETDB1 responsible for catalyzing the deposition of H3K9me3, a repressive histone mark promoting heterochromatin formation at targeted loci<sup>154</sup>. The two paralogs recognize, almost exclusively, overlapping sets of primate-restricted TEs, including members of the ERVK and ERV1 classes of ERVs as well as many SVA integrants<sup>154,156</sup>. To ask if the phenotype observed upon *ZNF587/417* KD cells correlated with changes in their chromatin status, we performed genome-wide profiling of H3K9me3 using CUT&Tag<sup>221</sup> in U2932 cells at day 3 of KD. Similar peak numbers ( $n=26,841$ ) and locations were detected in the control and KD cells, but in *ZNF587/417*-depleted cells peak intensity significantly decreased at 1,577 and increased at 755 locations, most of the latter in pericentromeric regions (89.9% vs. 10.07%,  $p < 0.0001$ ; Fig. 4a). Measurement of H3K9me3 across all called peaks confirmed a marked increase at pericentromeric regions and telomeres, contrasting with a global decrease across the rest of the genome following *ZNF587/417* KD (Fig. 4b).

While H3K9me3 peaks in control U2932 cells were enriched at TEs, simple repeats and satellite repeats, the KD triggered a preferential loss of this repressive mark at TEs and a gain at the other two categories of repeats (Fig. 4c). Amongst TEs, the most significant losers H3K9me3 upon *ZNF587/417* KD belonged to the ERV and SVA subgroups (Fig. 4d), notably ERVK integrants previously identified as the preferential ERV targets of these KZFPs<sup>154</sup> (Fig. 4e). When some increase in H3K9me3 was noted over TE inserts, it likely reflected spreading of this mark from surrounding satellite/simple repeats (Fig. 4c/f). Accordingly, peaks gaining H3K9me3 were on average substantially wider than TE-centered peaks losing this mark (Fig. 10a). Still, the loss of H3K9me3 over a range of TEs extending beyond the sole known targets of *ZNF587/417* suggested that additional factors contributed to this process. Confirming this suspicion, our RNA-seq analysis revealed the downregulation of another 100 *KZFP* genes ( $FDR < 0.05$ ) in U2932 cells 3 days after *ZNF587/417* KD (Fig. 4g), including 12 poor-prognosis *KZFP* gene hits. By contrast, only one *KZFP* gene was found upregulated in *ZNF587/417* KD cells at this time point. This downregulation was even broader at day 6 of



KD, encompassing 184 and 133 *KZFP* genes (FDR <0.05) besides *ZNF587* and *ZNF417* in U2932 and OCI-Ly7 cells, respectively (Fig. 10b). While the reason for such broad downregulation is unknown, we observed that the expression of *ZNF587/417* was highly correlated with that of other *KZFP* genes in our DLBCL cohort (Fig. 4h), pointing to common or cross-regulatory mechanisms within this gene family.

## ***ZNF587/417* depletion triggers replicative stress in lymphoma cells**

Our data pointed to a link between *ZNF587/417* depletion and a UV irradiation-like transcriptional response. To probe this issue further, we measured the impact of *ZNF587/417* KD on Ser-139 H2AX phosphorylation ( $\gamma$ H2AX) in U2932 and OCI-Ly7 cells 3 days after KD, which revealed a marked increase in this mark of DNA damage response (DDR) (Fig. 5a/11a). In order to determine whether this DDR was concentrated in particular regions of the genome, we performed CUT&Tag for  $\gamma$ H2AX in U2932 cells after 3 days of *ZNF587/417* KD. Due to the known critical role of peri-/centromeric regions and telomeres in chromosomal integrity, we used the new Telomere-to-Telomere assembly of the *CHM13* cell line (T2T-CHM13)<sup>222</sup> for our analysis. Peak numbers (n=7'635) and locations were similar in control and KD cells, but  $\gamma$ H2AX levels were significantly increased over 311 loci and decreased over 101 loci upon *ZNF587/417* KD (Fig. 5b), including 14 upregulated and 5 downregulated peaks located at pericentromeric regions (Fig. 5c). Although the number of peaks with a significant  $\gamma$ H2AX increase was not enriched at those regions, the peaks located at pericentromeric regions (n=480) and telomeres (n=42) presented a higher global fold increase in  $\gamma$ H2AX compared to other peaks upon *KZFP* KD (Fig. 5d). This reflected a significant enrichment in  $\gamma$ H2AX-gaining peaks overlapping with simple repeats compared with other  $\gamma$ H2AX-bearing loci (Fig. 5e). These results suggested that, when *ZNF587/417* are depleted, DNA damage was exacerbated at preexisting  $\gamma$ H2AX hotspots rather than at TE loci targeted by these KZFPs. Since exacerbated DDR at preexisting  $\gamma$ H2AX hotspots has been described with RS inducers such as hydroxyurea or aphidicolin<sup>223</sup>, we characterized DNA replication fork progression at the genome-wide level using Transferase-Activated End Ligation sequencing (TrAEL-seq). We observed that *ZNF587/417* KD cells exhibited a global loss of replication polarity, indicating a lower frequency of diverging replication forks related to an increase in the number of replication origins (Fig. 5f). The excessive firing of replication origins is a known source of RS, notably due to the exhaustion of limiting factors necessary for replication fork progression<sup>83,224</sup> and collisions between replication and transcription<sup>225</sup>, but it can also be a

protective mechanism from RS triggered by primary stalling of replication forks<sup>226</sup>. In support of the former hypothesis, we observed a strong correlation between H3K9me3 changes and replication fork polarity in *ZNF587/417* KD cells (Fig. 5g), arguing in favor of a link between these two alterations. Strikingly, genomic regions with a loss in H3K9me3 exhibited a decrease in replication directionality, whereas the opposite was observed where H3K9me3 was gained (Fig. 5h/11c), that is, predominantly at pericentromeric regions (Fig. 11b). To probe if the cell growth impairment observed upon *ZNF587/417* depletion was due to RS, we assessed DNA replication by measuring the incorporation of the thymidine analog 5-Ethynyl-2'-deoxyuridine (EdU) in control and knockdown U2932 and SUDHL4 cells by flow cytometry. It revealed a decrease in EdU incorporation, along with a proportional increase of non-replicating cells during the S phase of the cell cycle (Fig. 5i), thus confirming RS-related arrest.

### ***ZNF587/417* depletion leads to cell-intrinsic inflammation**

At day 6 of KD, 732 genes were upregulated and 915 downregulated (FC>2, FDR <0.05; Fig. 6a), including 195 the 332 (58.7%) genes already noted as differentially expressed at day 3 (Fig. 6b). The transcriptome was then characterized by the upregulation of IFN-responsive and other inflammatory genes, and by an enrichment in Hallmark gene sets previously noted in KZFP<sup>Low</sup> DLBCL samples (Fig. 12a). Interestingly, IFN-related DNA damage signature (IRDS) genes, such as *STAT1*, *IRF7*, *IFIT1/3*, *MX1*, *ISG15*, *IFI44* and *OASL* family members were upregulated at day 6 and 10, suggesting that the observed inflammatory response was at least partially related to genomic instability (Fig. 6c). TEs being well known sources of cell-intrinsic inflammation, we monitored their transcription during these events. Even though some TE loci were detected as mildly expressed at day 6, frank TE upregulation became apparent only at day 10 post-KD (Fig. 12b). Furthermore, the TEs significantly upregulated at day 6 belonged to the SINE group, probably induced by the ongoing genotoxic stress<sup>227</sup>. In contrast, most TEs significantly induced at day 10 were ERVs (Fig. 12c), notably HERV9, HERV9N, HERV9NC, and LTR12C subfamily members of the primate-specific ERV1 family harboring *ZNF587/417*-binding motifs (Fig. 12d). Of note, these ERV subsets already displayed an increased expression at day 6, albeit not reaching statistical significance at the TE family level (Fig. 12c). Irrespectively, our transcriptome analyses revealed that IFN/inflammatory-related genes were already upregulated at day 3 of *ZNF587/417* knockdown, that is, still in the complete absence of TE induction (Fig. 3c). Therefore, TE transcripts and their products could not be the source of the inflammation initially observed upon KZFP depletion.

The accumulation of cytosolic DNA molecules due to endogenous or exogenous genotoxic stress is another known trigger of an IFN response, occurring via activation of the cGAS-STING DNA sensing pathway<sup>228,121,229</sup> by micronuclei or free DNA fragments generated respectively in the context of lagging chromosomes<sup>126</sup> or replication defects<sup>129,229</sup>. At day 3 of *KZFP* KD, U2932 cells presented a significant accumulation of cytosolic dsDNA (Fig. 6d) as well as of the dsDNA sensing intermediate mediator phospho-STING (Fig. 6e), together with the phosphorylation and nuclear translocation of IRF3, one of its main downstream effectors (Fig. 6f). Accordingly, upregulation of IRF3 target genes was recorded at day 6 of *ZNF87/417* KD (Fig. 12e). Furthermore, transcripts encoding CXCL10, a chemokine promoting the attraction of mononuclear phagocytes, T cells, and Natural Killer (NK) cells, were upregulated starting from day 3 up to day 10 following *ZNF587/417* KD, and we could confirm increased levels of this cytokine in the cell supernatant (Fig. 12f). Induction of a cell-intrinsic inflammatory response was not restricted to ABC cells, as upregulation of IFN-related responsive genes similarly occurred in GCB-related OCI-Ly7 cells following *ZNF587/417* KD (Fig. 6g/12g), thus further highlighting that the effect of *KZFP* depletion was similar across different cellular contexts. Furthermore, upon intersecting genes differentially expressed at day 6 post-KD in U2932 or OCI-Ly7 cells with the genes defining the *KZFP*<sup>High</sup> and *KZFP*<sup>Low</sup> groups of the DLBCL cohort, we found that *ZNF587/417* KD induced a shift of the cell lines towards a *KZFP*<sup>Low</sup>-like transcriptome, supporting an important role for *ZNF417* and *ZNF587* in the gene expression profile of poor-prognosis DLBCL (Fig. 6h). In order to assess if these changes would lead to an increased susceptibility to innate immunity clearance, we tested the phagocytosis of human M1 polarized macrophages exposed to *ZNF587/417* KD cells during 24 hours between day 4 and 5 of KD (Fig. 6i). Indeed, M1 macrophages exhibited a stronger phagocytic ability against *ZNF587/417* KD cells compared to control cells, suggesting that in the clinics *ZNF587/417* blockade could foster the phagocytic clearance DLBCL residual disease. Interestingly, the transcriptome of the U2932 KD cells started reverting towards a *KZFP*<sup>High</sup> profile at day 10, despite a persistent *KZFP* downregulation, with re-induction of cell cycle-related genes such as E2F targets and G2M checkpoint genes, and downregulation of transcripts related to EMT, TNF $\alpha$  signaling via NF- $\kappa$ B, and hypoxia (Fig. 12h). This suggested ongoing transcriptional reprogramming, probably driven by inflammation and DNA repair, with progressive establishment of a resistant phenotype.

## ZNF587/417 depletion modifies the antigenic profile of lymphoma cells

Our data indicate that the RS induced by *ZNF587/417* KD triggers an inflammatory response through the cytosolic release of DNA fragments and their sensing by the cGAS-STING pathway. This type of response normally leads to increased expression of *HLA-I* genes, which play a pivotal role in the immunogenicity of cancer cells. Accordingly, in U2932 cells, *HLA-C* gene expression started to be significantly upregulated at day 3 of *ZNF587/417* KD (Fig. 9e), followed by *HLA-A*, *HLA-B*, and beta-2-microglobulin (*B2M*) at day 6 (Fig. 6c), while in OCI-Ly7 *HLA-B* and *HLA-C* transcripts were augmented only at this later time point (Fig. 12g). Corroborating these transcriptomic changes, U9932 cells displayed a 2-fold increase in surface expression of HLA-A, -B, and -C at KD day 10, even though at that point their source transcripts were no longer upregulated (Fig. 7a). We went on to assess whether the upregulation of HLA-I observed in *ZNF587/417* KD cells was associated with an alteration of their HLA-bound peptidome. For this, HLA-bound peptides of KD day 10 and control U2932 cells were eluted and sequenced by mass spectrometry. We found 1'367 peptides derived from 1050 different proteins (~14.3% of the detected immunopeptidome and 23% of presented proteins) to be significantly more and 1154 peptides from 876 proteins (~12% of the detected immunopeptidome and 19.2% of presented proteins) to be less abundant in *ZNF587/417* KD cells ( $P < 0.05$ ; Fig. 7b). In accordance with the transcriptome data, most of the peptides enriched in *ZNF587/417* KD cells corresponded to inflammatory gene products, while KZFP-derived peptides were prevalent amongst depleted entities (Fig. 7b). Furthermore, the peptide repertoire of U2932 KD cells was altered, with noticeable changes in peptide length: 8 and 9-mers made up 90.5% of the peptides enriched in KD cells but only 65.9% of the peptides enriched in control cells, and conversely 10- and 11-mers constituted only 8.8% of the peptides enriched in U2932 KD *versus* 29.4% in control cells (Fig. 7c). We hypothesized these changes to be linked to differences in HLA allele-specific expression. By using previously described HLA binding prediction methods (Fig. 7d), we found that the percentage of peptides predicted to bind HLA-B (69.4% *vs* 41.8%) and -C (7.7% *vs* 3.5%) was increased in U2932 KD cells compared to control cells, where peptides were predicted to bind mostly HLA-A03:01 (50.1%), an allele recently associated to poor responses to immune-checkpoint blockade<sup>230</sup>. The *ZNF587/417* KD increased also the proportion of hydrophobic (leucine) while decreasing that of amphipathic (lysine) residues at the C-terminus of the presented peptides (Fig. 7c),

suggesting underlying alterations of the proteasome degradation pathway and changes in HLA binding affinity. Therefore, we compared the immunogenicity of the eluted peptides using the Immune Epitope Database (IEDB) analytic tool<sup>231</sup>, which predicted that immunopeptides eluted from U2932 KD cells were on average more immunogenic than that of control cells (Fig. 7e). Together these results indicate that *ZNF587/417* KD reshuffles the immunopeptidome of lymphoma cells resulting in increased immunogenicity.

# DISCUSSION

Genomic instability confers cancer cells with evolutionary advantages by generating clonal diversity, which facilitates escape from immune surveillance and chemotherapeutic agents. Yet it has to be balanced in order to avoid mitotic catastrophes and anti-tumoral immune responses. Constitutive heterochromatin is a dynamic state reestablished after each cell division through epigenetic memory<sup>232,233,234</sup>. The present work demonstrates that KZFPs, the TE-targeting partners of the KAP1-SETDB1 complex, play an important role in the maintenance of heterochromatin in cancer cells, protecting these from excessive RS, secondary inflammatory responses and other immunologically detectable alterations. Our results further identify ZNF587 and ZNF417, two primate-specific paralogs, as leading pro-oncogenic KZFPs, and their upregulation as a strong negative prognosis indicator in DLBCL.

Recent studies have revealed that chromosomal instability (CIN) is a main driver of cancer progression and metastasis, and pointed to the induction of the cGAS-STING dsDNA sensing pathway as an important mediator of this process<sup>125</sup>. CIN<sup>High</sup> tumors are typically described as carrying a poor prognosis and displaying an activation of genes related to IFN/inflammatory responses, EMT and TNF $\alpha$  signaling<sup>125</sup>. Here, we found KZFP<sup>High</sup> DLBCLs to exhibit a higher average number of CNAs than their KZFP<sup>Low</sup> counterparts, yet their transcriptome to be characterized by an upregulation of DNA repair genes but CIN<sup>Low</sup> expression signatures linked to cell cycle, Myc target, and OxPhos-related genes. Thus, KZFP<sup>High</sup> DLBCL cells presumably combine higher degrees of genetic instability with an increased ability both to attenuate RS and to dampen inflammatory responses. As a consequence, the stabilization of their chromatin via upregulation of critical controllers of repetitive DNA results both in increasing the efficiency of their clonal expansion and diversification and in facilitating their immune evasion. ZNF587/417 recognize evolutionarily recent ERVs and SVAs, capable of robust transcriptional influences notably manifested during brain development<sup>156,154</sup>. In most tissues, *cis*-acting regulatory sequences embedded in these retroelements are maintained in a repressed state through a mix of histone and DNA methylation, the latter perpetuated across mitosis through the action of maintenance DNA methyltransferases<sup>30</sup>, without the need for sequence-specific repressors. Our data suggest a model whereby, perhaps due to the state of DNA hypomethylation characteristic of cancer cells, maintenance of heterochromatin at these TE loci becomes strongly dependent on high levels of histone methylation-inducing KZFPs recognizing their sequence in order to prevent inflammatory responses.

The importance of KZFPs in DLBCL pathogenesis is supported by the recurrence of amplifications of the distal portion of Chr19q in this lymphoid malignancy<sup>11,235</sup>. This phenomenon had been so far attributed to the presence in this chromosomal segment of *SPIB*, a gene encoding a repressor of the IFN $\beta$  gene activator IRF7<sup>11</sup>. However, *SPIB* acts in a COO-restricted manner<sup>236</sup>, whereas Chr19q amplifications have been detected in both ABC and GCB DLBCLs and demonstrated to correlate with poor prognosis independently from the lymphoma COO in prospective cohorts<sup>235</sup>. Our results thus strongly suggest that this chromosomal amplification is driven by the pro-oncogenic potential of *ZNF587/417* and other components of *KZFP* gene cluster 11, as well as perhaps by the additional presence in the sub-telomeric region of Chr19q of *KAP1*, which not only is an essential cofactor of these KZFPs but also itself a negative prognostic factor in non-Hodgkin B cell lymphoma<sup>237</sup>.

Our finding that gH2AX accumulates at hotspots of endogenous DNA damage upon *ZNF587/417* KD is reminiscent of chromatin alterations detected in lymphoblastoid cell lines exposed to chemical inducers of RS<sup>223</sup>. Moreover, the reshuffling of the immunopeptidome observed in lymphoma cells surviving the acute genotoxic stress triggered by *ZNF587/417* KD closely resembles changes observed upon exposure of tumor cells to non-cytotoxic doses of the nucleoside analogue gemcitabine<sup>238</sup>.

Together, these data indicate that methods aimed at inactivating *ZNF587/417* could mimic the action of chemotherapeutic agents. Our finding that normal cells do not undergo genotoxic stress when depleted in these KZFPs suggests that approaches targeting these TE controllers might have a very favorable therapeutic index, further warranting efforts geared toward their development.

# MATERIAL AND METHODS

**Cell culture.** OCI-Ly7 and U2932 human lymphoma cell lines were obtained from DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany ([www.dsmz.de](http://www.dsmz.de)). The other DLBCL cell lines, namely SUDHL4 and HBL1 were kindly supplied by the laboratory of Professor Elisa Oricchio (EPFL, Swiss Institute for Experimental Cancer Research) and typed using short tandem repeat (STR) profiling by Microsynth cell line authentication service. U2932, HBL1, and SUDHL4 cells were grown in RPMI 1640 (Gibco) with 10% fetal bovine serum and 1% penicillin-streptomycin. OCI-Ly7 cells were grown in IMDM (Gibco) with 20% fetal bovine serum (FBS) and 1% penicillin-streptomycin. All cell lines were grown at 37°C and 5% CO<sub>2</sub>. HEK293T cells were grown in DMEM (Gibco) supplemented with 10% FBS, 1% penicillin-streptomycin, and 2mM of L-glutamine. Primary normal dermal fibroblasts were cultured in Fibroblast Basal Medium (ATCC) supplemented with Fibroblast Growth Kit-Low Serum (ATCC). Human M1-polarized macrophages and M-CSF were purchased from StemCell. After thawing, M1 macrophages were resuspended in complete RPMI 1640 (Gibco) medium with M-CSF (StemCell) at a final concentration of 50 ng/ml.

**KZFP knockdown experiments.** The short hairpin sequences targeting ZNF587/417 were designed as described<sup>154</sup>. ZNF587/417 shRNAs are targeting the following sequences: 5'-GCAGCATATTGGAGAGAAATT-3' (shRNA.1) and 5'-AGTCGAAAGAGCAGCCTTATT-3' (shRNA.2). The other shRNAs were purchased from Sigma-Aldrich, with shRNAs against ZNF587B/814 and ZNF586 targeting the following sequences: 5'-CCTTCTAAGCAGAGTATTTAT-3' (shZNF587/814), 5'-GCTTATACATCTAGTCTCATT-3' (shZNF586). An shRNA with the non-targeting sequence 5'-CAACAAGATGAAGAGCACCAAG-3' (shScramble, shScr) was used as a negative control in all experiments. Lentivectors (LVs) were produced as described on the following website: <http://tronolab.epfl.ch>. Lymphoma cell lines and primary fibroblasts were transduced with a MOI of 10. 48h post-transduction, cells were selected for 3 days using 3 µg/mL of puromycin. For experiments performed between day 2 to 4, cells were used without selection while doubling the multiplicity of infection (MOI of 20).



**Cell proliferation assays.** The tetrazolium bromide (MTT) assay was used to determine cell proliferation for 4 days. Cells were plated after 5 days of LV transduction and 3 days of puromycin selection, except for SUDHL4 cells which were plated 3 days after LV transduction. Briefly, 100 $\mu$ L of cells were seeded in 96-well plates at a concentration of 5k/well for primary dermal fibroblasts, 20k/well for HBL1, and 40k/well for U2932, OCI-Ly7, and SUDHL4 cells. On each consecutive day, cells were incubated with 5 mg/ml MTT for 3h at 37°C. The reaction was stopped by the addition of 100 $\mu$ L of 10% Ultrapure SDS (Invitrogen) to dissolve precipitated crystals. After overnight incubation at 37°C, the absorbance was measured on a microplate spectrophotometer at 570 nm while subtracting the background absorbance measured at 690 nm.

**Annexin V/propidium iodide cell death detection assays.** Cell death was determined using flow cytometry through the quantification of cell surface annexin V-APC (Biolegend) and propidium iodide (PI, Biolegend) co-staining. Briefly, one million cells were harvested 3 days after transduction and washed twice with cell staining buffer (Biolegend). Cells were resuspended in 100 $\mu$ L of Cell Staining Buffer (Biolegend) and incubated in the dark for 15 minutes at room temperature following the addition of 5 $\mu$ L of Annexin V-APC antibody and 10 $\mu$ L of PI solution, following manufacturer's instructions. Samples were then analyzed on a BD LSR II (Becton Dickinson, USA) flow cytometer, using BD FACSDiva™ software, and quantified using FlowJo single-cell analysis software (FlowJo, LLC).

**Cell cycle analyses.** Cell cycle distribution was analyzed by flow cytometry measurement of cellular DNA content using PI staining. Three million cells were collected, washed, and resuspended in 1 volume of ice-cold 1x Dulbecco's Phosphate Buffer Saline (PBS, BioConcept) before being fixed by the addition of 2 volumes of ice-cold 100% ethanol during slow vortexing. After 45 min of incubation at 4°C, cells were washed again in ice-cold PBS and resuspended in 1.5 volume of ice-cold PI staining solution (0.1% triton, 200  $\mu$ g/ml RNase A, and 50  $\mu$ g/ml PI). After 30 minutes of incubation at room temperature in the dark, the samples were analyzed on a BD LSR II (Becton Dickinson, USA) flow cytometer, using BD FACSDiva™ software, and quantified using FlowJo single-cell analysis software (FlowJo, LLC).

***CXCL10 measurement in cell culture supernatant.*** CXCL10 concentration was determined using the Human ProQuantum Immunoassay Kit according to the manufacturer's instructions. This method relies on antigen recognition by two antibodies, each conjugated to a specific DNA oligo. The presence of the antigen will allow the proximity ligation of the DNA oligo from one antibody to the other, and further quantification of the ligation product by quantitative PCR, as a readout of CXCL10 concentration. Fresh supernatants from U2932 KD and control cells were harvested at day 6 post-transduction. Supernatants were mixed with the two oligonucleotide-conjugated antibodies, diluted 1:10 with assay dilution buffer, and incubated for 1h at room temperature. Oligonucleotide ligation was performed by the addition of Master mix and Ligase and incubation for 1h at room temperature. Quantification of the ligation products was determined by quantitative PCR on an Applied Biosystems 7900HT Fast Real-Time PCT system (ThermoScientific) using the recommended instrument settings. The Ct values were exported to the ProQuantum Cloud application ([apps.thermofisher.com/apps/proquantum](https://apps.thermofisher.com/apps/proquantum)) to determine CXCL10 concentrations. Finally, supernatant protein concentrations were normalized by cell concentration measured at the beginning of the experiment.

***Analysis of published datasets.*** The BAM files from 775 exon-enriched DLBCL transcriptomes were downloaded from the European Genome-phenome Archive (EGA) at the European Bioinformatics Institute under the accession number EGAS00001002606<sup>208</sup>. To ensure better analysis quality, we excluded 142 samples with fewer than 1 million counts in total. After filtering lowly expressed genes with an average expression <1 CPM, 23'560 genes were subjected to a univariate Cox regression analysis using the R-package "survival". For this purpose, each gene expression in normalized counts units was compared to the overall survival (OS) followed by the application of an independent Cox model for each gene. 5'220 genes were found to be significantly associated to survival (Wald test *p-value* <0.05) and 1'530 remained after correction for multiple testing using the Benjamini-Hochberg method (FDR <0.05). KZFP genes were identified using a de novo generated whole genome annotation (Supplementary Table 1) using Hidden Markov Model (HMM) profile generated from seed sequences downloaded from Pfam<sup>239</sup> and created using hmmbuild from HMMER2 ([hmmer.org](http://hmmer.org)). KZFP genes were further identified as protein-coding using Ensembl BioMart database (GRCh38.p13, v103)<sup>240</sup>. Other human TF genes were identified using the previously published list from Lambert et al<sup>138</sup>. The categorization of the KZFPs was based on their evolutionary age as previously reported by Imbeault et al<sup>156</sup>.

Clinical data (OS, IPI details, and response to first-line therapy), driver mutation status, and COO classification of DLBCLs were extracted from the metadata provided in the publication of Reddy et al.<sup>208</sup>. The relationship between the 41 KZFP genes identified in Cox regression analysis was measured by Spearman rank correlation using the rcorr function of the Hmisc R-package and plotted using pheatmap. *KZFP* genes genomic cluster affiliation and evolutionary ages were extracted from Pontis et al.<sup>152</sup> and Imbeault et al.<sup>156</sup>, respectively. Hierarchical clustering analysis of lymphoma samples based on the expression of the 18 KZFPs located in the Chr19q co-expression module was computed using default parameters of the heatmap.2 function from gplots using Euclidean distances and Ward's agglomeration method. The clusterProfiler R-package was used for enrichment analysis of DEGs using Hallmarks gene sets<sup>212</sup>. Copy number alterations were inferred from VCF files provided by Reddy et al. with gains and losses called as described in their publication<sup>208</sup>.

**Data analysis.** Unless otherwise specified, graphs were obtained by GraphPad Prism software version 9 or ggplot2 R package.

**RT-qPCR.** Total RNA extraction was performed using the NucleoSpin RNA plus kit (Machery-Nagel) according to the manufacturer's recommendations. cDNA synthesis for qPCR was conducted using the Maxima H minus cDNA synthesis master mix (ThermoScientific). Real-time quantitative PCR was performed using PowerUp SYBR Green Master Mix (ThermoScientific) and run on a QuantStudio™ 6 Flex Real-Time PCR System.

**RNA-seq libraries and downstream analyses.** RNA-seq libraries were prepared using the Illumina Truseq Stranded mRNA kit. Libraries were sequenced in 75 or 100 bp paired-end formats on the Illumina Hiseq 4000 and NovaSeq 6000 sequencers, respectively. RNA-seq reads were mapped to the hg19 and T2T-CHM13v2.0 human genome releases using hisat v2.1.0<sup>241</sup>. Only uniquely mapped reads were used for counting over genes and repetitive sequence integrants (MAPQ > 10). Counts for genes and TEs were generated using featureCounts v2 and normalized for sequencing depth using the TMM method implemented in the limma package of Bioconductor<sup>242</sup>. Counts on genes were used as library size to correct both gene and TE expression. For repetitive DNA elements, an in-house curated version of the Repbase database was used. Details about the creation of this list can be found in this reference<sup>154</sup>. Differential gene expression analysis was performed using Voom<sup>243</sup> as

implemented in the Limma package of Bioconductor<sup>244</sup>. P-values were adjusted for multiple testing using the Benjamini-Hochberg method.

***Cleavage Under Targets and Tagmentation (CUT&Tag).*** CUT&Tag was performed as described in Kaya-Okur et al<sup>221</sup> without modifications. For each mark, 150k cells were used per sample using the anti-H3K9me3 primary antibody (Active Motif, AB\_2532132), the anti- $\gamma$ H2AX primary antibody (Abcam, ab2893), and anti-rabbit IgG (Abcam, ab46540) secondary antibody. A homemade purified pA-Tn5 protein (3XFlag-pA-Tn5-Fl, addgene #124601) was produced and coupled with MEDS Oligos by the Protein Production and Purification of EPFL, as previously described<sup>245</sup>. Purified recombinant protein was used at a final concentration of 700 ng/uL (1:250 dilution from homemade stock). Libraries were sequenced with 75 bp paired end on the NextSeq 500 (Illumina). Reads were aligned to the T2T-CHM13v2.0 reference genome<sup>222</sup> using bowtie2<sup>246</sup>. Only proper read pairs with MAPQ>10 were kept. CUT&Tag peaks were called using SEACR v1.3<sup>247</sup> on “stringent” mode and numeric threshold 0.01 for H3K9me3 and 0.001 for  $\gamma$ H2AX, and merged with maximum distance allowed of 2000bp for H3K9me3 and 0bp for  $\gamma$ H2AX with bedtools 2.27.1. Encode blacklist for hg19 genome<sup>248, 249</sup> was lifted over to T2T-CHM13v2.0 genome and filtered out from the annotated peaks. Bedtools multicov was used to count mapped reads on the annotated and filtered peaks, and differential peak analysis was performed using Voom after library size correction (using the total number of aligned reads as size factor) performed using the TMM method. Bigwig coverage tracks with the mean of replicate samples were generated using bedtools 2.27.1<sup>250</sup> and deeptools 3.3.1<sup>251</sup>, and heatmap representations of the coverage signal were performed using computeMatrix function and plotHeatmap from deeptools 3.3.1.

***EdU DNA synthesis monitoring flow cytometry.*** Lymphoma cells were pulse-labeled with 10  $\mu$ M 5-ethynyl-2'-deoxyuridine (EdU) for 20 min and subsequently fixed with 2% formaldehyde for 30 min at room temperature (RT). EdU incorporation was detected using Click chemistry according to manufacturer's instructions (Click-iT EdU Flow Cytometry Cell Proliferation Assay, Invitrogen). Cells were resuspended in 1x Phosphate Buffer Solution (PBS, BioConcept) with 1% bovine serum albumin (BSA), 2  $\mu$ g/ml DAPI and 0.5 mg/ml RNase A for 30 min at RT and subsequently analyzed on a BD LSR II (Becton Dickinson, USA) flow cytometer, using BD FACSDivaTM software, and quantified using FlowJo single-cell analysis software (FlowJo, LLC).

**TrAEL-Seq.** TrAEL-Seq was performed as described in the publication of Kara et al<sup>252</sup>. In Brief, 1 million of cells transduced with ShScr and ShRNA.1 were collected 3 days after LV transduction. Cells were washed once in 10 ml L buffer (10 mM Tris HCl (pH 7.5), 100 mM EDTA, 20 mM NaCl) and resuspended in 60 µl L buffer in a 2-ml tube. Samples were heated to 50°C for 2 to 3 min before the addition of 40 µl molten CleanCut agarose (Bio-Rad 1703594), vortexed vigorously for 5 sec before pipetting in plug mould (Bio-Rad 1703713), and solidified at 4°C for 30 min. Each plug was transferred to a 2-ml tube containing 500 µl digestion buffer (10 mM Tris HCl (pH 7.5), 100 mM EDTA, 20 mM NaCl, 1% sodium N-lauroyl sarcosine, 0.1 mg/ml Proteinase K) and incubated overnight at 50°C. Plugs were washed and treated with RNase T1. Half of an agarose plug was used for each library. For restriction enzyme digestion, a plug was equilibrated 30 min in 200 µl 1x CutSmart buffer (NEB), digested overnight at 37°C with 1 µl 20 U/µl NotI-HF (NEB R3189S) and 1 µl 10 U/µl PmeI (NEB R0560S) in 400 µl 1x CutSmart buffer, then 1 µl 20 U/µl SfiI (NEB R0123S) was added and incubation continued overnight at 50°C. The plug was rinsed with 1x TE before further processing. Plugs were equilibrated once in 100 µl 1x TdT buffer (NEB) for 30 min at room temperature, then incubated for 2 h at 37°C in 100 µl 1x TdT buffer containing 4 µl 10 mM ATP and 1 µl Terminal Transferase (NEB M0315L). Plugs were rinsed with 1 ml Tris buffer (10 mM Tris HCl (pH 8.0)), equilibrated in 100 µl 1x T4 RNA ligase buffer (NEB) containing 40 µl 50% PEG 8000 for 1 h at room temperature, then incubated overnight at 25°C in 100 µl 1x T4 RNA ligase buffer (NEB) containing 40 µl 50% PEG 8000, 1 µl 10 pM/µl and previously adenylated using the 5' DNA adenylation kit (NEB, E2610S), and 1 µl T4 RNA ligase 2 truncated KQ (NEB M0373L). Plugs were then rinsed with 1 ml Tris buffer, transferred to 15 ml tubes, and washed 3 times in 10 ml Tris buffer with rocking at room temperature for 1 to 2 h each, then washed again overnight under the same conditions. Plugs were equilibrated for 15 min with 1 ml agarase buffer (10 mM Bis-Tris-HCl, 1 mM EDTA (pH 6.5)), then the supernatant removed and 50 µl agarase buffer added. Plugs were melted for 20 min at 65°C, transferred for 5 min to a heating block preheated to 42°C, 1 µl β-agarase (NEB M0392S) was added and mixed by flicking without allowing sample to cool, and incubation continued at 42°C for 1 h. DNA was ethanol precipitated with 25 µl 10 M NH<sub>4</sub>OAc, 1 µl GlycoBlue, 330 µl of ethanol and resuspended in 10 µl 0.1x TE. A volume of 40 µl reaction mix containing 5 µl isothermal amplification buffer (NEB), 3 µl 100 mM MgSO<sub>4</sub>, 2 µl 10 mM dNTPs, and 1 µl Bst 2 WarmStart DNA polymerase (NEB M0538S) was added and sample incubated 30 min at 65°C before precipitation with 12.5 µl 10 M NH<sub>4</sub>OAc, 1 µl GlycoBlue, 160 µl ethanol and redissolving pellet in 130 µl 1x TE. The DNA was transferred to an AFA microTUBE (Covaris

520045) and fragmented in a Covaris E220 using duty factor 10, PIP 175, Cycles 200, Temp 11°C, then transferred to a 1.5-ml tube containing 8 µl prewashed Dynabeads MyOne streptavidin C1 beads (Thermo, 65001) resuspended in 300 µl 2x TN (10 mM Tris (pH 8), 2 M NaCl) along with 170 µl water (total volume 600 µl) and incubated 30 min at room temperature on a rotating wheel. Beads were washed once with 500 µl 5 mM Tris (pH 8), 0.5 mM EDTA, 1 M NaCl, 5 min on wheel and once with 500 µl 0.1x TE, 5 min on wheel before resuspension in 25 µl 0.1x TE. TrAEL-seq adaptor 2 was added using a modified NEBNext Ultra II DNA kit (NEB E7645S): 3.5 µl NEBNext Ultra II End Prep buffer, 1 µl 1 ng/µl sonicated salmon sperm DNA (this is used as a carrier), and 1.5 µl NEBNext Ultra II End Prep enzyme were added and reaction incubated 30 min at room temperature and 30 min at 65°C. After cooling, 1.25 µl 10 pM/µl TrAEL-seq adaptor 2, 0.5 µl NEBNext ligation enhancer, and 15 µl NEBNext Ultra II ligation mix were added and incubated 30 min at room temperature. The reaction mix was removed and discarded and beads were rinsed with 500 µl wash buffer (5 mM Tris (pH 8), 0.5 mM EDTA, 1 M NaCl), then washed twice with 1 ml wash buffer for 10 min on wheel at room temperature and once for 10 min with 1 ml 0.1x TE. Libraries were eluted from beads with 11 µl 1x TE and 1.5 µl USER enzyme (NEB) for 15 min at 37°C, then again with 10.5 µl 1x TE and 1.5 µl USER enzyme (NEB) for 15 min at 37°C, and the 2 eluates combined. Amplification of the library was performed with components of the NEBNext Ultra II DNA kit (NEB E7645S) and a NEBNext Multiplex Oligos set (e.g., NEB E7335S). An initial test amplification was used to determine the optimal cycle number for each library. For this, 1.25 µl library was amplified in 10 µl total volume with 0.4 µl each of the NEBNext Universal and any NEBNext Index primers with 5 µl NEBNext Ultra II Q5 PCR master mix. Cycling program: 98°C 30 s, then 18 cycles of (98°C 10 s, 65°C 75 s), 65°C 5 min. Test PCR was cleaned with 8 µl AMPure XP beads (Beckman A63881) and eluted with 2.5 µl 0.1x TE, of which 1 µl was examined on a Bioanalyser high sensitivity DNA chip (Agilent 5067–4626). Ideal cycle number should bring final library to final concentration of 1 to 3 nM, noting that the final library will be 2 to 3 cycles more concentrated than the test anyway. A volume of 21 µl of library was then amplified with 2 µl each of NEBNext Universal and chosen Index primer and 25 µl NEBNext Ultra II Q5 PCR master mix using same conditions as above for calculated cycle number. Amplified library was cleaned with 40 µl AMPure XP beads (Beckman A63881) and eluted with 26 µl 0.1x TE, then 25 µl of this was again purified with 20 µl AMPure XP beads and eluted with 11 µl 0.1x TE. Final libraries were quality controlled and quantified using a Bioanalyzer system (2100 Agilent 5067–4626) and KAPA qPCR (Roche KK4835).

Preparation of TrAEL-seq adaptors: DNA oligonucleotides were synthesized and PAGE purified by Sigma-Genosys (Merck, United Kingdom). Sequences were as follows:

1:[Phos]NNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTUGCGCAGGCCATTGGCC[BtdT]GCGCUACACTCTTCCCTACACGACGCT.

2:[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCUUUUGACTGGAGTTCAGACGTGTGCTCTTCCGATC\*T.

Libraries were sequenced on an Illumina NextSeq 500 sequencer as 75 bp paired-end reads. All scripts used for the analyses of TrAEL-seq data can be found in the following git repository: <https://github.com/fmartins/traelseq>). The paired-end TrAEL-seq reads carry an 8-bp in-line barcode (UMI) at the 5'-end of the first read in the pair, followed by a variable number of 1 to 3 thymines (T). Read structure of the first read in the pair is, therefore, NNNNNNNN(T)<sub>n</sub>SEQUENCESPECIFIC, where NNNNNNNN is the UMI, and(T)<sub>n</sub> is the poly(T). We used flexbar<sup>253</sup> to 1) remove the first 8 bp (UMI) of each read, 2) for each read in the pair, add the UMI sequence to the end of the readID, and 3) remove up to 3 T (inclusive) at the start of the sequence. UMI processed files were then aligned to the T2T-CHM13v2.0 genome using Bowtie2<sup>246</sup> (parameters: --local -D 15 -R 2 -N 1 -L 25 -i S,1,0.75). Alignment files (bams) were then deduplicated using umitools<sup>254</sup> (parameters: --paired --extract-umi-method read\_id --method unique). Reads were then summed in running windows using the bamCoverage function from deepTools<sup>251</sup>. Count tables per region in the format of bedgraph files were then imported into R<sup>255</sup> for plotting and statistics. Bigwig files were computed using the bedGraphToBigWig software<sup>256</sup>. For read polarity plots, the following formula was used: read polarity = (R - F)/(R + F), where F and R relate to the total forward and reverse read counts respectively. To compute R and F, the following parameters were used in bamCoverage: for forward (--samFlagInclude 64 --samFlagExclude 16) and for reverse (--samFlagInclude 80).

**Immunofluorescence.** U2932 cells were fixed and permeabilized using True-Nuclear™ Transcription Factor Buffer Set (Biolegend) following manufacturer's instructions, incubated with primary anti-dsDNA antibody (1:100, Mybiosource), anti-phospho-STING Ser365 (1:400, CST), or anti-phospho-IRF3 Ser396 (1:100, CST) for 30 minutes, washed, and incubated with secondary anti-Mouse IgG Alexa Fluor 488 (1:100, Invitrogen) or anti-Rabbit IgG Alexa Fluor 568 (1:100, Invitrogen) for another 30 minutes. Nuclei were stained with DAPI (4', 6-diamidino-2-phenylindole, 1 µg/mL, Sigma-Aldrich). Images were acquired on a point-scanning confocal microscope SP8 Leica with an oil 40x objective 1.25 N.A. Images

are Z-stacks of dimensions 256x256 in XY and spanning the entire cell in Z (voxel size 120x120x500 nm). Pinhole was set at 1 Airy Unit for an emission wavelength of 520 nm for all channels, acquired in sequential scan mode. DAPI was excited with a 405 nm laser (emission 420-470nm), its emission collected between 420 and 470 nm. The excitation of the DAPI channel is used to make a transmission image. DsDNA (Alexa 488) was excited with a 488 nm laser, its emission collected between 500 and 530nm. Phospho-STING and phospho-IRF3 (Alexa 568) were excited with a 552 nm laser, their emission was collected between 560 and 610 nm. DsDNA quantification estimation was performed in ImageJ/Fiji<sup>257</sup>. Each image is projected in Z using maximum intensity projection. The DAPI channel was subsequently median filtered (radius = 2 pixels), and the nuclear region was delineated by using a fixed intensity threshold, identical for all images. A band of 0.8 micrometer surrounding the nucleus is then defined by shrinking the nucleus region by 5 pixels, then expanding it by 0.8 micrometer, thus defining a cytosolic band region. Average intensities of DAPI and dsDNA channels were then measured within the shrank nuclear region and the cytosolic band. All experiments were reproduced at least once.

**Phagocytosis assay.** Human M1 macrophages (effector cells) were plated into 96-well flat clear bottom black walled culture plates (ThermoScientific 165305) at a density of  $1 \times 10^4$ - $2.5 \times 10^4$  cells/well in 50 uL complete RPMI medium supplemented with M-CSF (50 ng/ml) overnight. U2932 cells (target cells) were transduced as described above, harvested at day 4, washed with pHrodo Wash bBffer (Sartorius), and labeled using pHrodo red (Sartorius) at a final concentration of 31.3 ng/mL for 45 minutes at 37°C in pHrodo Labeling Buffer (Sartorius), before being washed again and resuspended in complete RPMI medium.  $3 \times 10^4$ - $7.5 \times 10^4$  target cells/well were then seeded over effector cells at target:effector ratio of 3:1. 96-well plates were then placed into the InCucyte S3 System (Sartorius) for time-lapse imaging. The pHrodo signal was measured by the acquisition of images from 3-5 fields/well every 1 to 1:30 hours for 24 hours.

**HLA typing and prediction of HLA binding affinity.** HLA typing using HLA-HD v1.4.0<sup>258</sup> on RNA sequencing data from U2932. For each sample, the consensus HLA haplotypes were determined as the consensus typing derived from all RNA sequencing replicates. NetMHCpan-4.1 tool was used for predicting the binding affinity of the eluted peptides to the respective HLA alleles<sup>259</sup>.



**Immunopeptidomics analysis.** HLA-I peptides were immunopurified from U2932 cells. A first set of control and *ZNF587/417* KD cell pellets (30 million cells each) was used to generate a spectral library from data-dependent and data-independent acquisition methods (DDA and DIA, respectively). A second set of pellets (with at least 20 millions of cells) were acquired using DIA method and analyzed by Spectronaut. The two sets of samples were processed as previously described to purify the HLA-I peptides<sup>260</sup>. Briefly, anti-pan HLA-I monoclonal antibodies (W6/32) were purified from supernatant of HB95 hybridoma cells (ATCC® HB-95) using protein-A sepharose 4B (Pro-A) beads (Invitrogen) and cross-linked to pro-A beads. Each sample was lysed in 1.5mL lysis buffer comprising of PBS containing 0.25% sodium deoxycholate (Sigma-Aldrich), 0.2 mM iodoacetamide (Sigma-Aldrich), 1 mM EDTA, 1:200 Protease Inhibitors Cocktail (Sigma-Aldrich), 1 mM Phenylmethylsulfonylfluoride (Roche, Basel, Switzerland), 1% octyl-beta-D glucopyranoside (Sigma-Aldrich), at 4°C for 1 hour and centrifuged at 4°C at 14200 rpm for 50 min using a refrigerated table-top centrifuge (Eppendorf Centrifuge, Hamburg, Germany). The cleared lysates were loaded on wells of 96-well single-use filter micro-plates with 3 µm glass fibers and 25 µm polyethylene membranes (Agilent, 204495-100) containing the HB95 cross-linked beads. Once the lysates flew through, the plates were washed several times with solutions containing various concentrations of salts and two final washes of 2mL of 20mM Tris-HCl pH8 using the Waters Positive Pressure-96 Processor (Waters). The HLA complexes bound to the beads were then directly eluted into preconditioned Sep-Pak tC18 100 mg Sorbent 96-well plates (Waters, 186002321) using 1% trifluoroacetic acid (TFA; Sigma Aldrich). After two washes of 1mL 0.1% TFA, the HLA-I peptides were eluted using 25% acetonitrile (ACN; Sigma Aldrich) in 0.1% TFA. Peptides were dried using vacuum centrifugation (Concentrator plus, Eppendorf) and stored at -20 °C. Prior to LC-MS/MS analysis, iRT peptides (Biognosys) were spiked in each sample for retention time calibration through Spectronaut analysis. For each sample, dried peptides were re-suspend in 12 ul of which 3ul samples were injected for DDA and 2 ul for DIA.

The Easy-nLC 1200 liquid chromatograph (Thermo Fisher Scientific) was coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) for peptide tandem mass spectrometry data analysis. Peptides were separated on a home-made 50cm analytical column with a tip size of ~10µm (OD = 365µm, ID = 75µm) and packed with ReproSil-Pur C18 (1.9-µm particles, 120 Å pore size, Dr Maisch GmbH). The peptide separation by chromatography was performed as previously described<sup>261</sup>. For MS/MS data acquisition in DDA and DIA, the same ion sampling methods described by Pak et al.<sup>262</sup> were used.

**DDA and DIA data analysis.** We used Spectronaut 15.7.220308.50606<sup>263</sup> to build a spectral library from DDA and DIA MS/M data. The peptide identification search was done by Pulsar<sup>TM</sup> using an unspecific digestion type and a peptide length from 8 -15 a.a. with a maximum variable modification of 5 per peptide. Acetylation of n-term and oxidation of methionine were used as variable modifications. The library was then generated using a peptide and PSM FDR  $\leq 0.01$  and protein FDR = 1 from Pulsar search. The DIA MS/MS data was matched against the library as previously described by Pak et al.<sup>262</sup>, with the exception that this time iRT peptide calibration was used.

**Data Availability.** The accession number for the RNA-seq and CUT&Tag data generated in this article is GEO:

# ACKNOWLEDGMENTS

We thank C. Raclot, P.-Y. Helleboid, J. Duc, V. Glutz, R. Guiet, N. Chiaruttini and J. Pontis for their technical support and scientific advice; R. Guindon for help in graphical output of the figures, the EPFL Flow Cytometry (FCCF), Protein Production and Purification (PTPSP) and Gene Expression (GECF) core facilities for the use of their systems and services. We thank A. Mayran and the Laboratory of Developmental Genetics of Prof. Denis Duboule for help with the use of their InCucyte S3 Sartorius imaging system. We also acknowledge all the members of the Trono laboratory for fruitful discussions and constant support. This study was supported by grants from the Swiss National Science Foundation and the European Research Council (KRABnKAP, no. 268721; Transpos-X, no. 694658), the Personalized Health and Related Technologies (PHRT) strategic focus area of the Swiss Federal Institutes of Technology (ETH) Domain and the Swiss Personalized Health Network (SPHN) initiative of the Swiss Academy of Medical Sciences (project no. 2017-407) and by the generous longstanding support of the Aclon Foundation to D.T.

## AUTHOR CONTRIBUTIONS

F.M. and D.T. conceived the study and designed experiments; F.M., O.R., and R.F. performed the wet experiments with the technical support of S.O.; J.C., E.P., C.P., and F.M. conducted the bioinformatics analyses, and F.M., O.R., R.F., J.CF., and D.T. wrote the manuscript, with review and corrections by all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the lead contact, Didier Trono.

# FIGURES AND LEGENDS

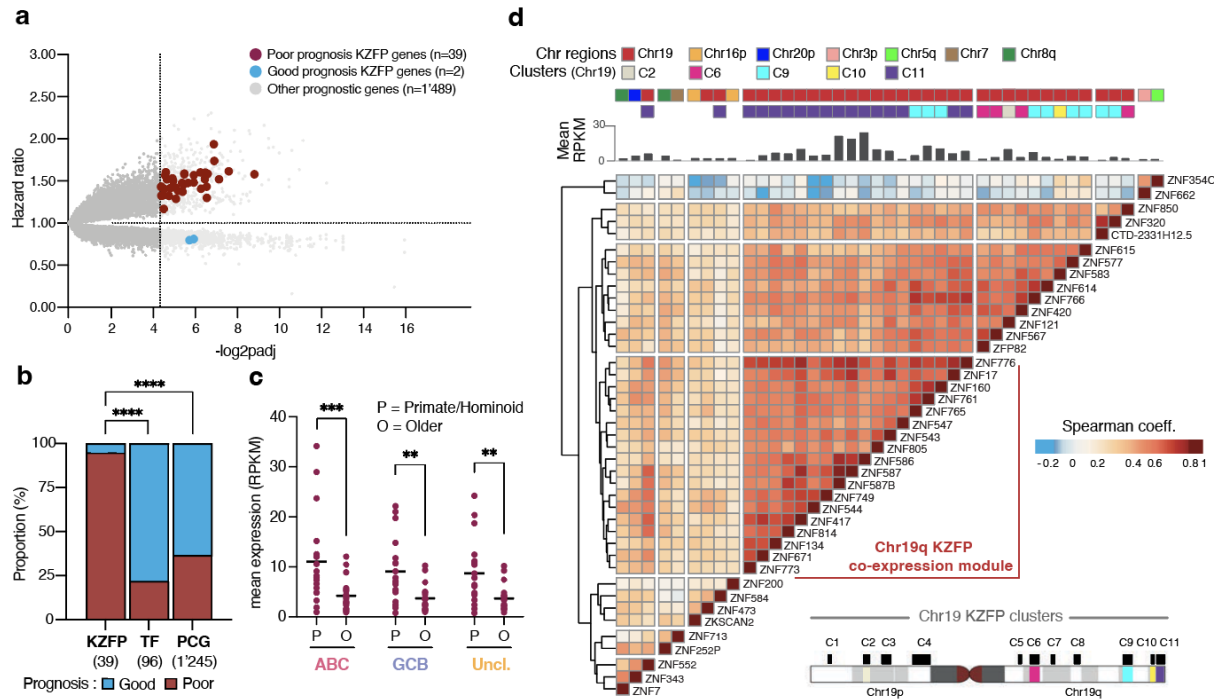
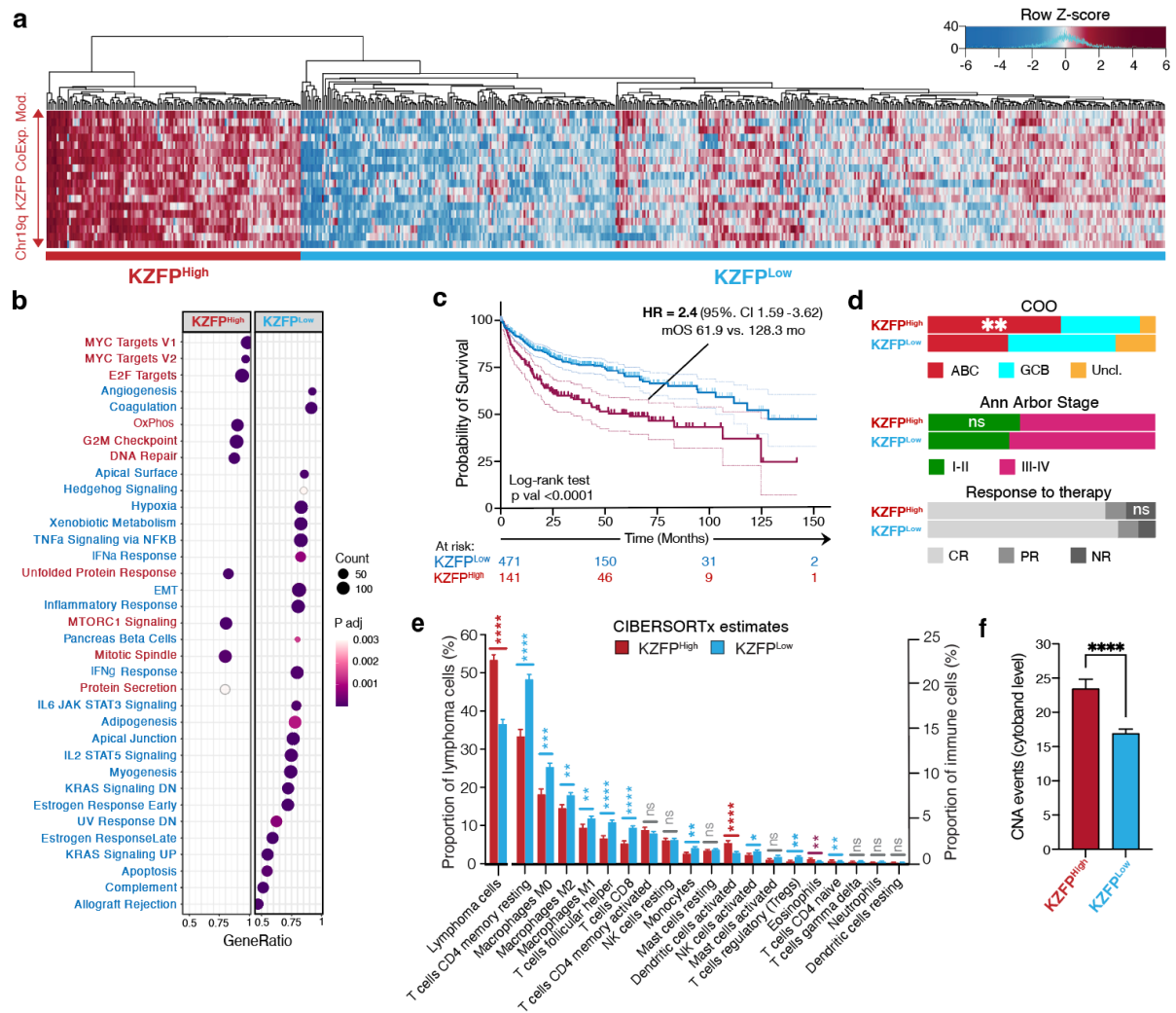


Figure 1 | A KZFP gene cluster as poor prognosis predictor in DLBCL.

(a) Volcano plot of hazard ratios (HRs) issued from univariate Cox regression analysis correlating individual gene expression with patient survival outcome are plotted against significance (n=612 DLBCL RNA-seq with available survival data). The vertical dashed line indicates the threshold of significance (FDR < 0.05). The horizontal dashed line separates genes associated with an increased (>1) and reduced risk (<1) of death. (b) Stacked bar plots showing the proportion of gene hits associated with good (blue) or poor prognosis (dark red). Genes were hierarchically categorized as protein-coding KZFPs, other TFs, and other protein-coding genes (PCGs). Statistics: Two-tailed Fisher's Exact Test. (c) Dot plot showing the mean expression in reads per kilobase of exon per million reads mapped (RPKM) of the 19 primate/hominoid-specific (P) and 20 evolutionary older (O) protein-coding KZFP gene hits correlated with survival across ABC (n=256), GCB (n=281) and Unclassifiable (n=96) DLBCL samples. Statistics: Two-sided Mann–Whitney U-test. (d) Hierarchical clustering and heatmap showing the spearman correlation coefficients of pairwise comparison between the 41 KZFP gene hits based on their gene expression across 633 DLBCL samples – Euclidean distances and complete-linkage clustering method. The bar graph on top of the heatmap shows the mean expression in RPKM and the upper annotation track shows the chromosomal location. The

lower right ideogram displays KZFP gene hits positioning onto previously described Chr19q KZFP clusters<sup>152</sup>.

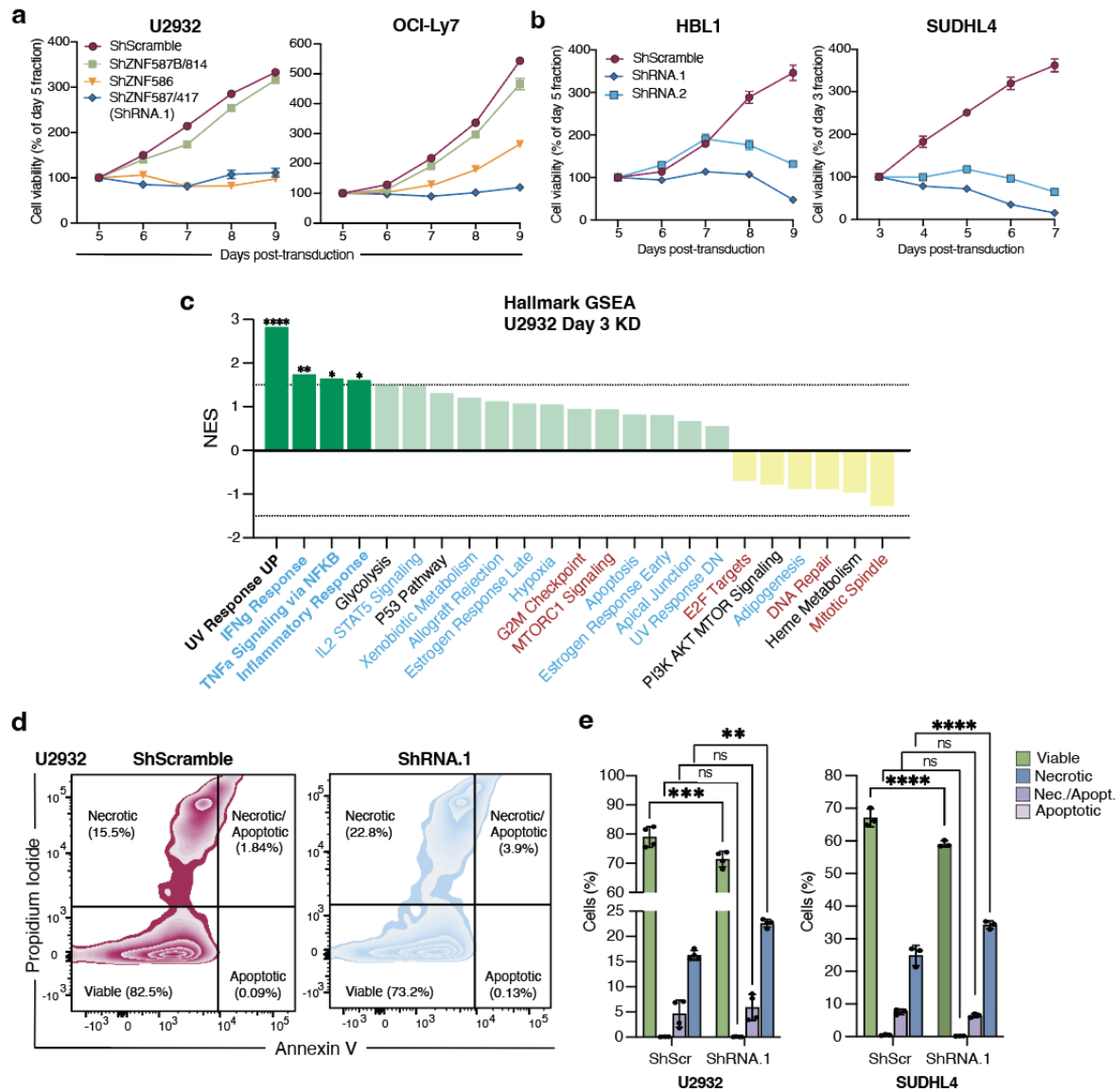
Statistical significance is indicated by the asterisks as follows: \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ ; \*\*\*\*,  $p < 0.0001$ .



**Figure 2 | KZFP<sup>High</sup> DLBCLs display cell-autonomous growth features.**

(a) Heatmap of the unsupervised hierarchical clustering analysis of 18 KZFP genes expression of the Chr19q co-expression module across 633 DLBCL samples – Euclidean distances and Ward’s agglomeration method. (b) Dot plots of enriched Hallmark gene sets in KZFP<sup>High</sup> vs. KZFP<sup>Low</sup> DLBCLs. The X-axis represents the ratio of differentially expressed genes (DEGs) in these groups over the total number of genes of each pathways (gene ratio). Dot size represent the number of DEGs in each gene set (count) and color scale represents the False Discovery Rate (FDR < 0.05). (c) Kaplan-Meier survival curves of patients displaying a KZFP<sup>High</sup> vs. KZFP<sup>Low</sup> DLBCL signature as defined in (a). Statistics: Log-rank test. (d) Stacked bar plots showing the proportion of the 3 different COO subtypes, the Ann Arbor stages at diagnosis, and the response to initial rituximab-based standard regimen amongst the patients belonging to each KZFP-defined group. CR = complete response; PR = partial response; NR = no response. Statistics: Two-tailed Fisher’s Exact Test. (e) Bar plot showing the differences in the mean

proportion (+/- SEM) of lymphoma cells (left) and immune cell infiltrate (right) between KZFP<sup>Low</sup> and KZFP<sup>High</sup> DLBCLs using CIBERSORTx RNA-seq bulk deconvolution. Statistics: Two-sided Mann–Whitney U-test with Benjamini &Yekutieli correction. (f) Bar plots showing the mean number (+/-SEM) of copy number alterations (CNA) detected by WES conducted on the same samples included in the transcriptomic analysis.

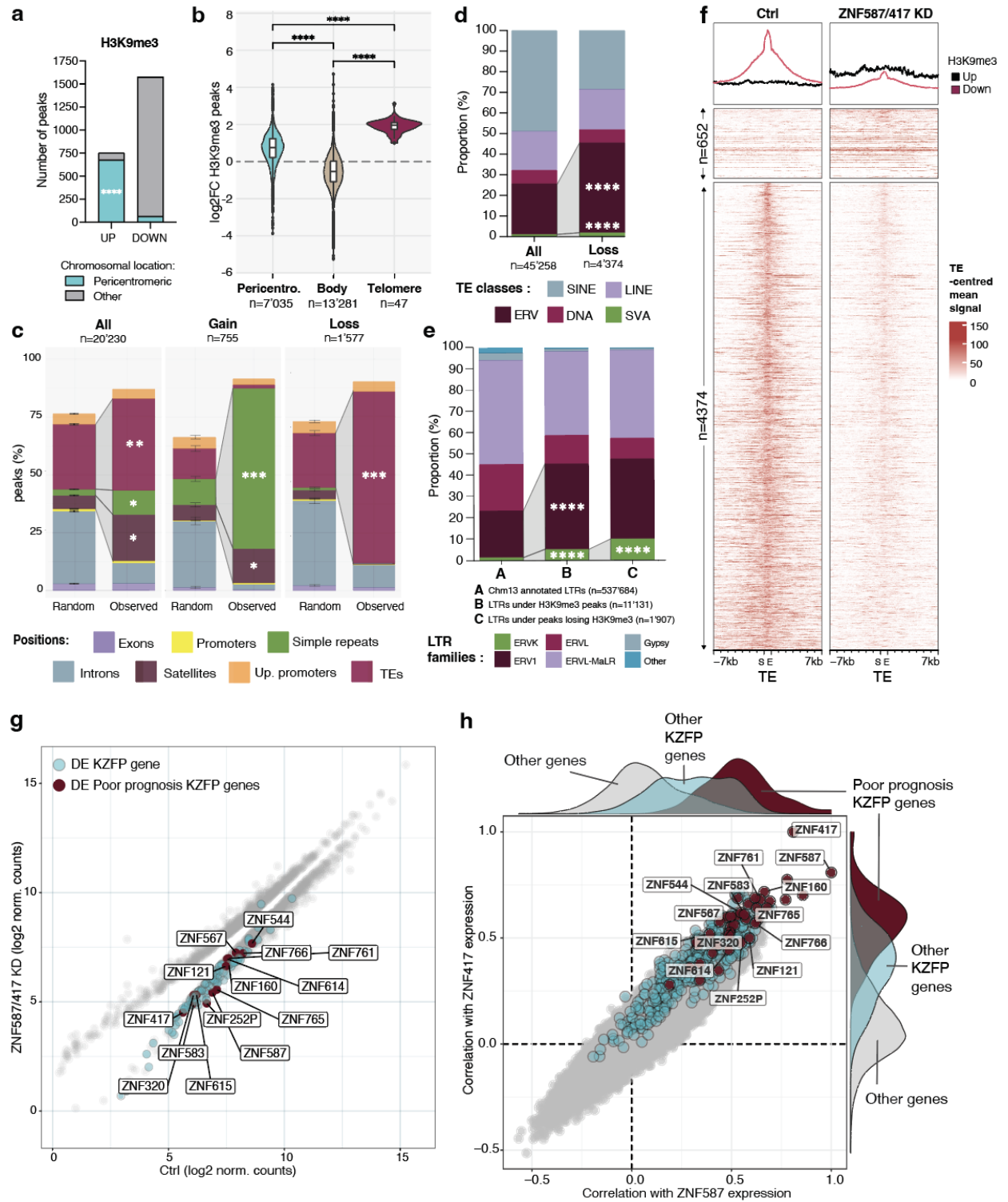


**Figure 3 | *ZNF587/417* depletion impairs cell growth and viability.**

(a) MTT proliferation assays of U2932 and OCI-Ly7 cells upon LV transduction with shRNAs targeting *ZNF587B/814*, *ZNF587/417*, and *ZNF586* or a control shRNA (shScr). (b) MTT proliferation assays for HBL1 and SUDHL4 upon LV transduction with 2 different shRNAs targeting *ZNF587/417* (shRNA.1/2) or a control shRNA. (c) Waterfall plot of Hallmark GSEA signatures from RNA-seq data ranked by their Normalized Enrichment Score (NES)<sup>236</sup>. Enriched signatures (NES>0) are highlighted in green and depleted signatures (NES<0) are in yellow. The p-value cutoff <0.05 is indicated by a dotted line. Hallmark gene sets enriched in KZFP<sup>High</sup> or KZFP<sup>Low</sup> DLBCLs are highlighted in dark red and blue, respectively. (d) Representative flow cytometry plot using Annexin V-APC/PI staining for cell death



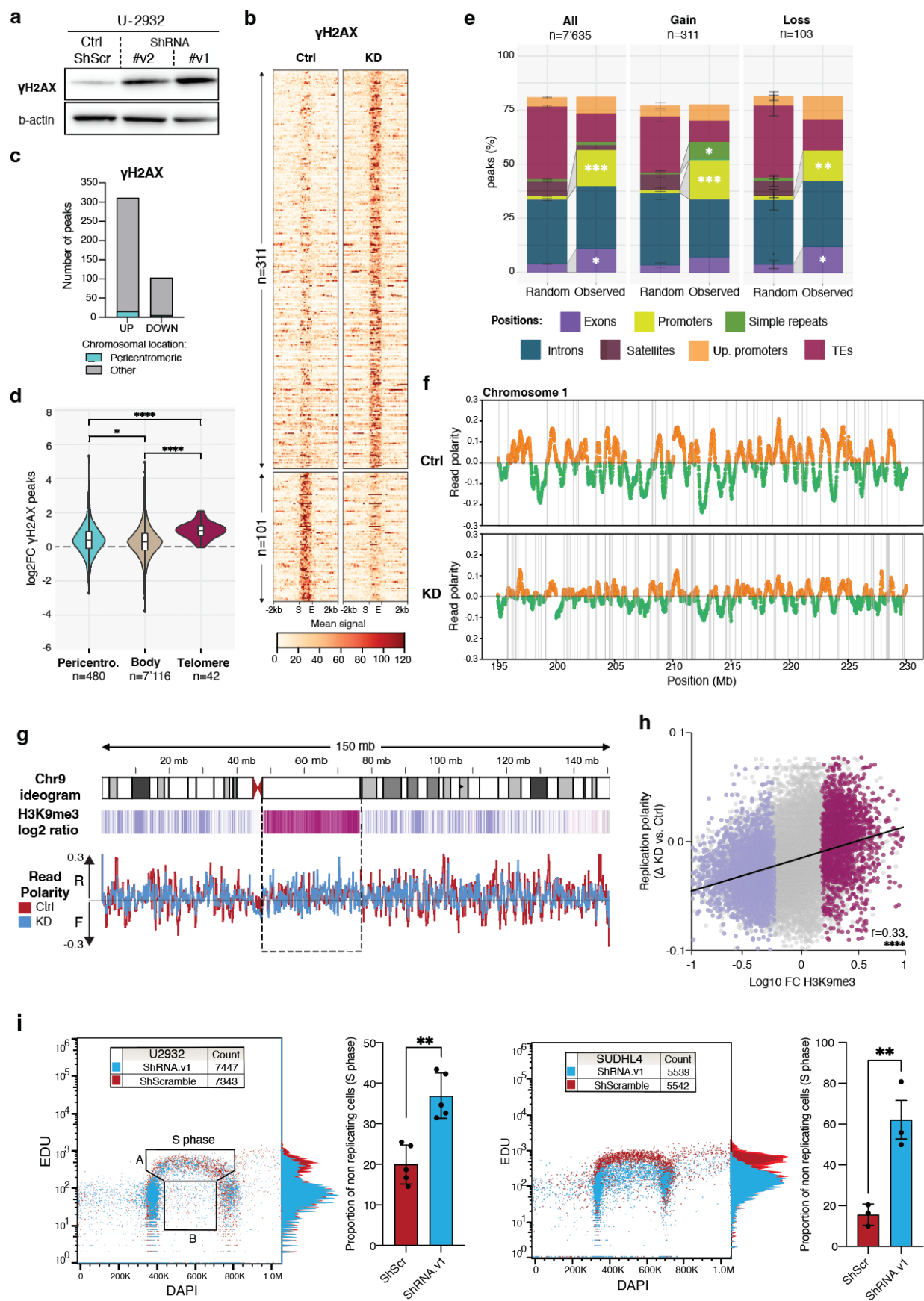
characterization in U2932 cells. (e) Bar plots showing the percentage of cells undergoing early (PI-/aV+) vs. late apoptosis (PI+/aV+) and necrosis (PI+/aV-) in U2932 and SUDHL4 cells, 3 days post-transduction with shRNA.1 targeting *ZNF587/417* or a control shRNA (mean  $\pm$  SD of independent duplicates). Statistics: Two-Way ANOVA with Sidak's multiple comparison test.



**Figure 4 | *ZNF587/417* depletion alters the heterochromatin landscape.**

(a) Number of Cut&Tag peaks with significant (FDR <0.05) gain (UP) or loss (DOWN) of H3K9me3, 3 days post-transduction with shRNA.1 targeting *ZNF587/417* or a control shRNA (two independent duplicates). (b) Violin plots of the distribution of the log<sub>2</sub> fold change (shRNA.1 vs. control) of H3K9me3 signal over called peaks in the listed genomic features. Statistics: Two-sided Mann–Whitney U-test. (c) Observed genomic distribution of all

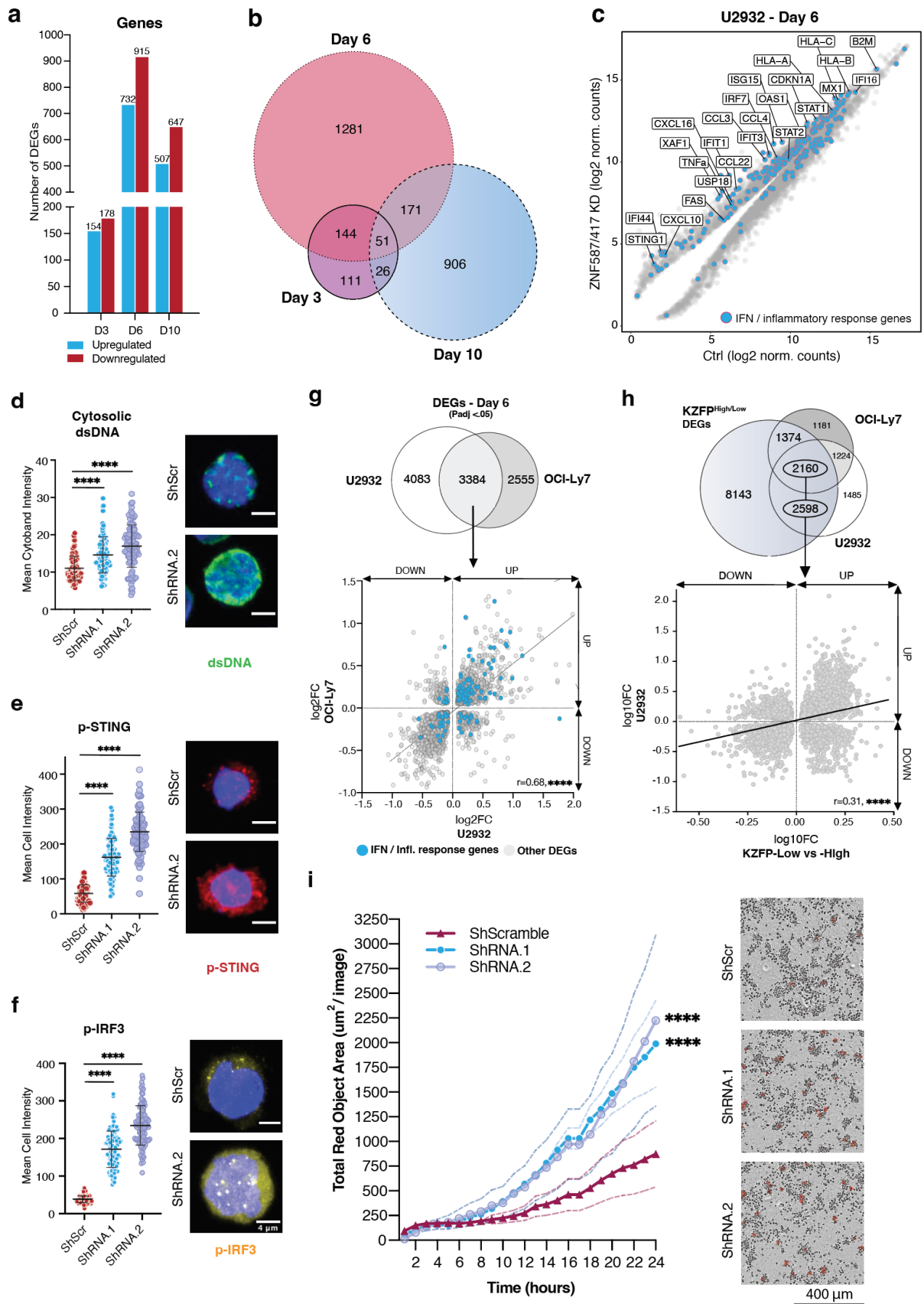
H3K9me3 peaks (left) and peaks displaying a significant gain (center) or loss (right) of H3K9me3 signal vs. randomized peaks, over listed genomic features (mean value of duplicates  $\pm$  SD from 1000 permutations; FDR <0.05). Genomic positions overlapping with >50% an H3K9me3 peak were assigned in the following order: gene promoters (up to 1000bp upstream from the transcription start site, TSS), up.promoters (up to 5000bp upstream from TSS), exons, TEs, satellites, simple repeats, and introns. Statistics: Chi-squared test, alternative: greater (d) Proportion of TE classes overlapping all H3K9me3 called peaks (left) or peaks with a significant loss of H3K9me3 (FDR <0.05, right). Asterisks indicate a significant enrichment compared to all H3K9me3 called peaks. Statistics: Two-tailed Fisher's Exact Test. (e) Same as (d) over LTR families, including peaks annotated in the T2T-CHM13v2.0 genome release (left). Statistics: Two-tailed Fisher's Exact Test (f) Profile heatmaps centered on TEs ( $\pm$  7kb) displaying significant change in H3K9me3 Cut&Tag signal (>50% overlap ; mean signal coverage from duplicates). (g) Scatter plot of RNA-seq data from ZNF587/417 KD versus control cells at day 3 of KD in U2932, outlining DEGs (grey dots, FDR <0.05), poor-prognosis KZFP genes (dark red dots) and other KZFP genes (turquoise dots). (h) Scatter plot of Spearman's correlation coefficients between ZNF417 (y-axis) or ZNF587 (x-axis) expression, and each gene expressed amongst the 633 DLBCLs analyzed transcriptomes (n= 23'560), outlining poor-prognosis KZFP genes (dark red dots), other KZFP genes (turquoise dots), and other genes (grey dots), with density plots showing their distribution on each side.



**Figure 5 | *ZNF587/417* depletion triggers replicative stress.**

(a) Western blot analysis of  $\gamma$ H2AX in U2932 cells 3 days after transduction with 2 different shRNAs targeting *ZNF587/417* (shRNA.v1/.v2) or a control shRNA (shScr).  $\beta$ -actin was used as a loading control. (b) Profile heatmaps of significantly changing  $\gamma$ H2AX Cut&Tag peaks (p-val. <0.05) in U2932 control and KD conditions, 3 days post-transduction (mean signal coverage from duplicates). (c) Number peaks with a significant gain (UP) or loss (DOWN) in  $\gamma$ H2AX in U2932 KD cells vs. control (p val <0.05). (d) Violin plots of the distribution of the log2 fold change (shRNA.1 vs. control) of  $\gamma$ H2AX signal over called peaks in the listed genomic features. Statistics: Two-sided Mann–Whitney U-test. (e) Observed genomic distribution of all  $\gamma$ H2AX peaks (left) and peaks displaying a significant gain (center) or loss (right) of  $\gamma$ H2AX signal, vs. randomized peaks, over listed genomic features (mean value of duplicates  $\pm$  SD from 1000 permutations; FDR <0.05). Genomic positions overlapping with >50% a  $\gamma$ H2AX peak were assigned in the following order: gene promoters (up to 1000bp upstream from the transcription start site, TSS), up.promoters (up to 5000bp upstream from TSS), exons, TEs, satellites, simple repeats, and introns. Statistics: Chi-squared test, alternative: greater. (f) TrAEL-seq read polarity plots over the chromosome 1 in U2932 control and KD conditions, 3 days post-transduction. Read polarity was calculated using a 250 kb sliding window spaced every 10 kb across the genome – *see methods*. Orange and green points represent sliding windows with a reverse or forward strand bias, respectively. Vertical grey lines represent shifts from forward to reverse bias, highlighting strong replication origins. (g) Top: Heatmap track of log2 ratio of Cut&Tag H3K9me3 signal between control and KD conditions surrounding the pericentromeric region of the chromosome 9. Significant H3K9me3 gain is highlighted by the dashed rectangle. Bottom: Overlapped read polarity plots of control (dark red) and KD (blue) conditions. (h) Scatter plot of replication polarity vs. H3K9me3 Cut&Tag signal changes. X-axis: log10 fold change of all H3K9me3 peaks called in U2932 cells. Y-axis: Replication polarity difference between KD and control conditions calculated from the absolute value of the percentage of reverse reads in a 2 Mb window surrounding the center of each H3K9me3 peak. Peaks with a significant gain in H3K9me3 are shown in dark red, the one with a significant H3K9me loss in turquoise blue, and others in grey. (i) Representative flow cytometry analysis of the replication signal (EdU incorporation intensity, y-axis) and DNA content (DAPI staining, x-axis) in U2932 and SUDHL4 KD (blue) vs. control cells (dark red), 3 days post-transduction. Y-axis side histogram shows changes in EdU intensity between KD (blue) and control (dark red) conditions. S phase cells are defined as

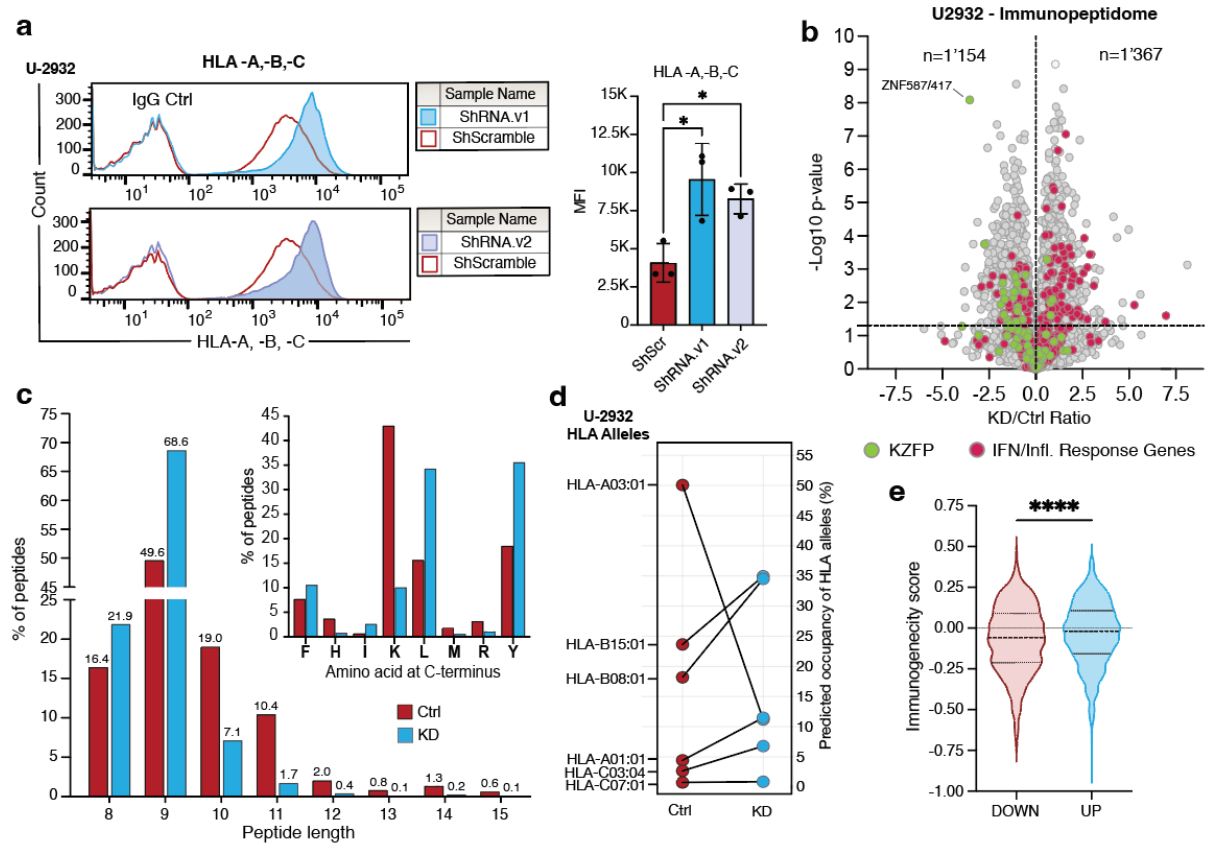
events located within the A+B gates ; A area= replicating cells, B area = non-replicating cells. Right: Barplots depicting the mean proportion ( $\pm$ SD) of non-replicating S phase cells calculated as follows:  $(B/(A+B)) * 100$  from 5 and 3 biological replicates in U2932 and SUDHL4, respectively. Statistics: Student's t-test.



**Figure 6 | ZNF587/417 depletion leads to cell-intrinsic inflammation.**

(a) Number of DEGs of RNA-seq analysis performed at day 3 (n=2), day 6 (n=3), and day 10 (n=2) of ZNF587/417 KD vs control U2932 cells. (b) Euler diagrams of the overlap of DEGs upon ZNF587/417 KD in U2932 cells at each time point (day 3, 6, and 10). (c) Scatter plot of log<sub>2</sub> normalized counts of ZNF417/587 KD cells vs. control U2932 at day 6, outlining DEGs (grey dots) among which genes belonging to type I/II IFN and Inflammatory response Hallmark gene sets, Interferon Signaling Reactome, and cellular response to type I, II and III IFN gene ontology terms are depicted in blue. (d) Representative images and dot plot showing the mean intensity of cytosolic double stranded DNA per cell, measured on z-stack immunofluorescence images (n ≥ 80 cells per condition). Statistics: Two-sided Mann–Whitney U-test. (e,f) same as (d) for phosphorylated STING (e) and phosphorylated IRF3 (f) signal respectively. (g) Top: Venn diagrams showing the overlap of DEGs upon ZNF587/417 KD in U2932 and OCI-Ly7 at day 6. Bottom: Scatterplot of Log<sub>2</sub> fold changes of DEGs shared between U2932 (x-axis) and OCI-Ly7 (y-axis) KD cells. Blue dots highlight the genes belonging to IFN-/Inflammatory response terms detailed in (c). The best-fit line in grey results from the linear regression analysis of U2932 log<sub>2</sub> fold changes onto OCI-Ly7 log<sub>2</sub> fold changes. (h) Top: Venn diagrams showing the overlap of DEGs upon ZNF587/417 KD in U2932 and OCI-Ly7 cells at day 6 and genes discriminating KZFP<sup>High</sup> and KZFP<sup>Low</sup> DLBCLs. Bottom: Scatterplot of log<sub>2</sub> fold changes of DEGs shared between genes discriminating KZFP<sup>High</sup>/KZFP<sup>Low</sup> DLBCLs (x-axis) and U2932 KD cells (y-axis). The best-fit line in black results from the linear regression analysis of KZFP<sup>High</sup> vs KZFP<sup>Low</sup> log<sub>2</sub> fold changes onto U2932 KD log<sub>2</sub> fold changes. For this panel, DEGs were defined with a FDR < 0.05 and a fold change > 2. r: correlation coefficient. (i) Phagocytosis assay of U2932 pHrodo-labeled cells co-cultured with M1-polarized macrophages. Representative 24h course of pHrodo signal quantification using Incucyte time-lapse imaging of ZNF587/417 KD cells with two different shRNAs (shRNA.1 in blue and shRNA.2 in turquoise blue) and control cells (shScramble in dark red). Total pHrodo cell area per image acquired was calculated for each time-point and plotted as a time course for each condition. Right: representative images of respective conditions. Statistics: Two-way ANOVA followed by Bonferroni correction. Asterisks indicate significant differences at 24h time-point compared to controls.

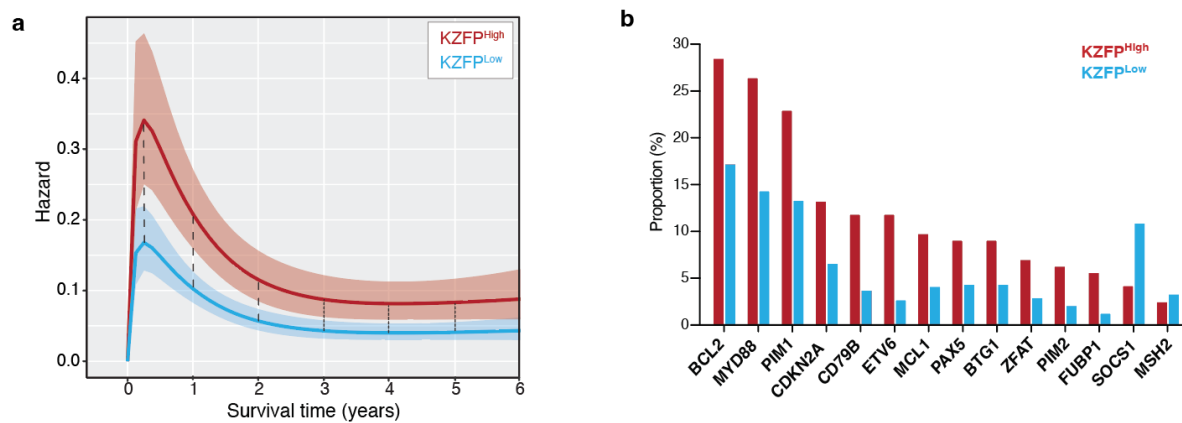




**Figure 7 | *ZNF587/417* depletion modifies the antigenic profile of lymphoma cells.**

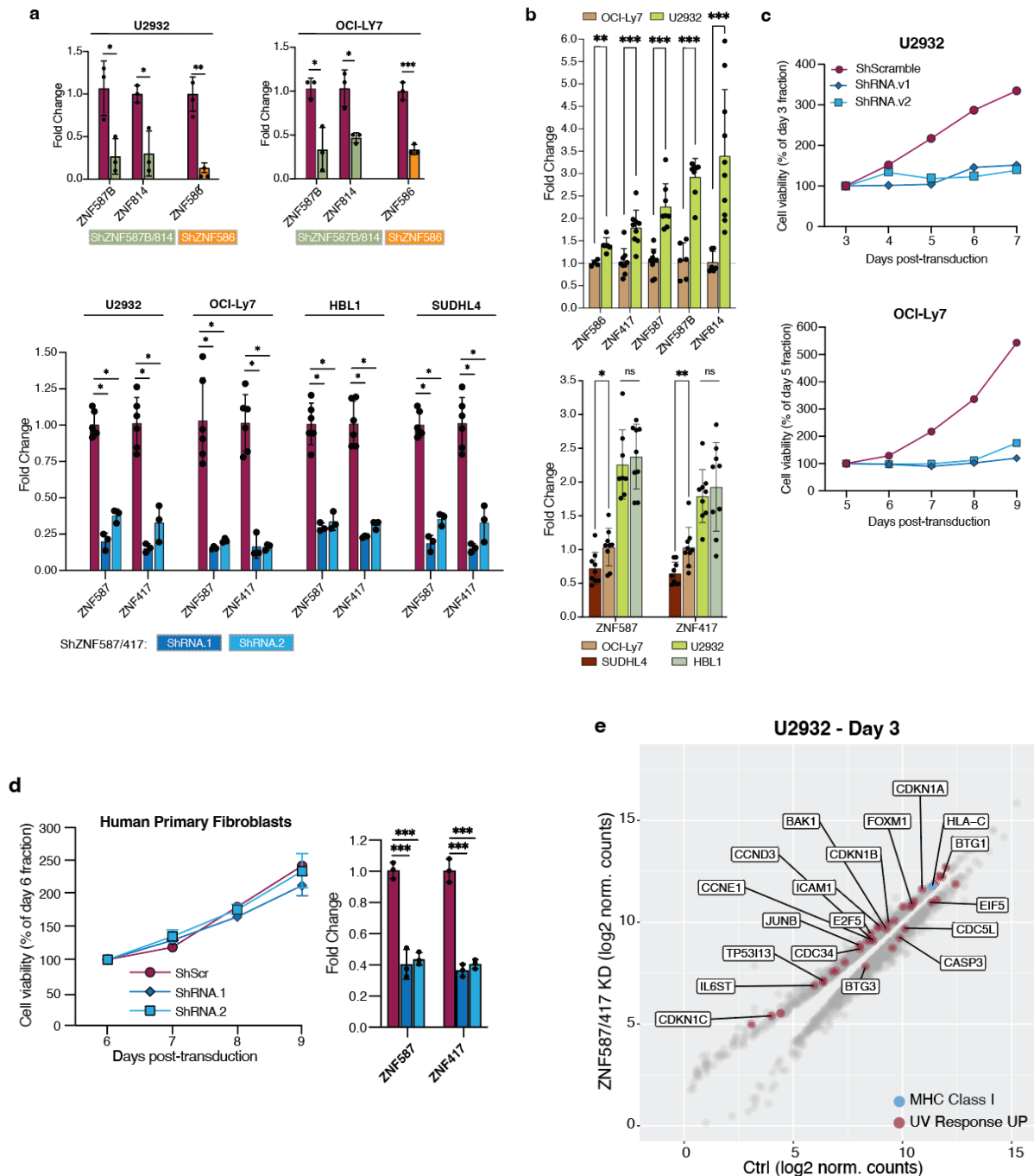
(a) Representative flow cytometry histograms (left) and bar plot quantification (right) of median fluorescence intensity (MFI) of HLA class I (HLA-A, -B, -C) expression on U2932 KD and control cells after 10 days with two different shRNAs targeting *ZNF587/417* (shRNA.1/2) or a control shRNA (shScr). Statistics: Student's t-test (mean of MFI  $\pm$  SD of independent triplicates). (b) Volcano plot showing changes in the abundance of HLA-I-bound peptides between *ZNF587/417* KD (shRNA.1/2) and control cells from 3 independent experiments. The horizontal dashed line indicates a p-value threshold of 0.05. n, number of enriched and depleted peptides. Peptides derived from gene products belonging to type I/II IFN and Inflammatory response Hallmark gene sets, Interferon Signaling Reactome, and cellular response to type I, II, and III IFN gene ontology terms are highlighted in dark red. Peptides derived from KZFPs are highlighted in green. Statistics: Student's t-test. (c) Bottom left: Differences in the proportion of peptides of different lengths between control (dark red) and KD cells (blue). Top right: Bar plot showing the differences in the percentage of peptides with listed C-terminal amino acid between control (dark red) and KD cells (blue). (d) Line plot showing the changes in the proportion of predicted occupancy of HLA-alleles present in U2932

cells between control (dark red) and KD cells (blue). (e) Violin plots showing the predicted immunogenicity scores of eluted peptide ligands enriched in control (DOWN, red) and KD cells (UP, blue) from the IEDB analytic tool.



**Figure 8 | Figure S1 related to Figure 2: A module of KZFP genes in DLBCLs with cell-autonomous growth features.**

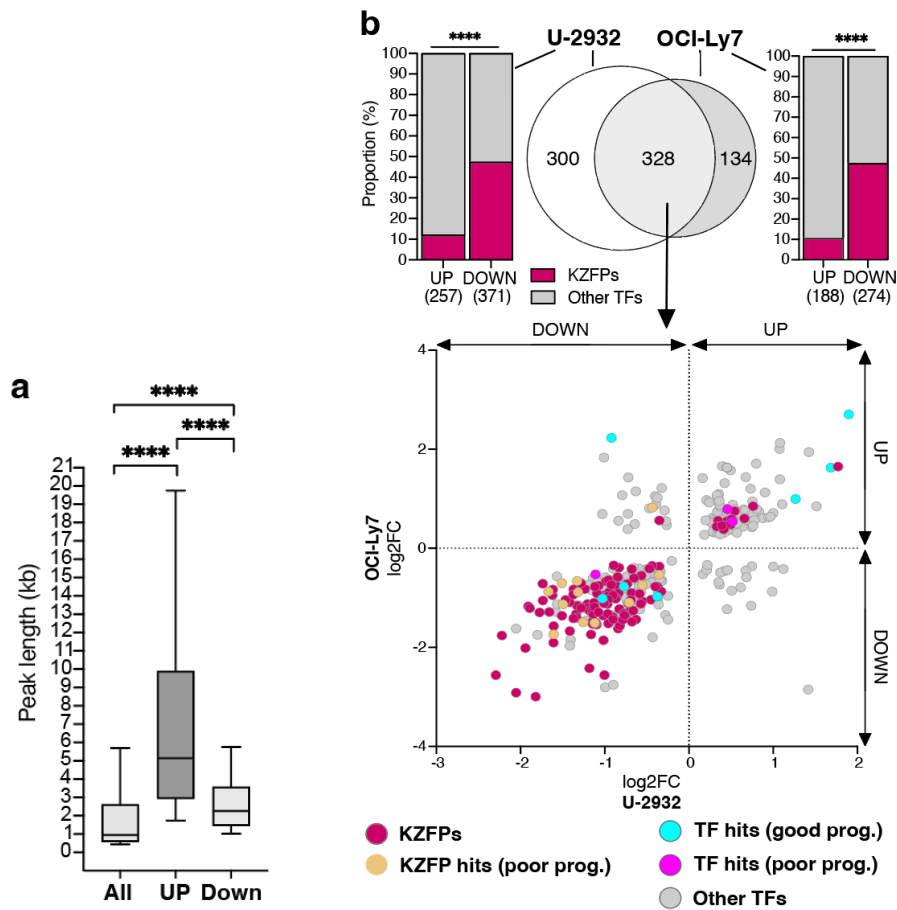
(a) Smoothed hazard functions by KZFP<sup>High/Low</sup> grouping for death among patients treated with a standard rituximab-based regimen, with, in dark red the hazards of patients diagnosed with a KZFP<sup>High</sup> and in blue with a KZFP<sup>Low</sup> DLBCL. Vertical dashed lines indicate the difference in hazard between the two groups at given time points after diagnosis. (b) Bar plots depicting the proportion of samples with given driver mutations in KZFP<sup>High</sup> (dark red) and KZFP<sup>Low</sup> (blue) groups. Only significantly enriched or depleted (FDR <0.05) driver mutations are depicted. Statistics: Two-tailed Fisher's Exact Test.



**Figure 9 | Figure S2 related to Figure 3: ZNF587/417 depletion impairs lymphoma cell growth and viability.**

(a) Quantitative RT-PCR analysis of ZNF587B, ZNF814, ZNF587, ZNF417, and ZNF586 mRNAs 5 days after LV transduction with shRNAs targeting ZNF586, ZNF587B/ZNF814, and ZNF587/ZNF417 paralog pairs. Statistics: Student's t-test. (b) Top: Bar plots showing the relative expression of ZNF586, ZNF417, ZNF587, ZNF587B, and ZNF814 in U2932 compared to OCI-Ly7 using GUSB, TBP, and/or ALAS2 housekeeping genes for

normalization. Bottom: Bar plots showing the relative expression of ZNF587 and ZNF417 in SUDHL4, U2932, HBL1 cells compared to OCI-Ly7 using GUSB, TBP, and/or ALAS2 housekeeping genes for normalization. (c) MTT proliferation assays of U2932 and OCI-Ly7 upon LV transduction with two different anti-ZNF587/417 shRNAs, or control shRNA (shScr). Cells were plated after 5 days of LV transduction and 3 days of puromycin selection. (d) MTT proliferation assay of human primary dermal fibroblasts upon LV transduction with two anti-ZNF587/417 shRNAs or control shRNA. Cells were plated after 5 days of LV transduction and 3 days of puromycin selection. Quantitative RT-PCR analysis of ZNF587 and ZNF417 mRNAs 5 days after LV transduction. Statistics: Student's t-test. (e) Scatter plot of RNA-seq from ZNF417/587 U2932 KD versus shScr control cells at day 3 of KD, outlining DEGs (grey dots, FDR <0.05) and among them genes belonging to the UV response UP Hallmark gene set (dark red dots) and MHC Class I genes (blue dots).

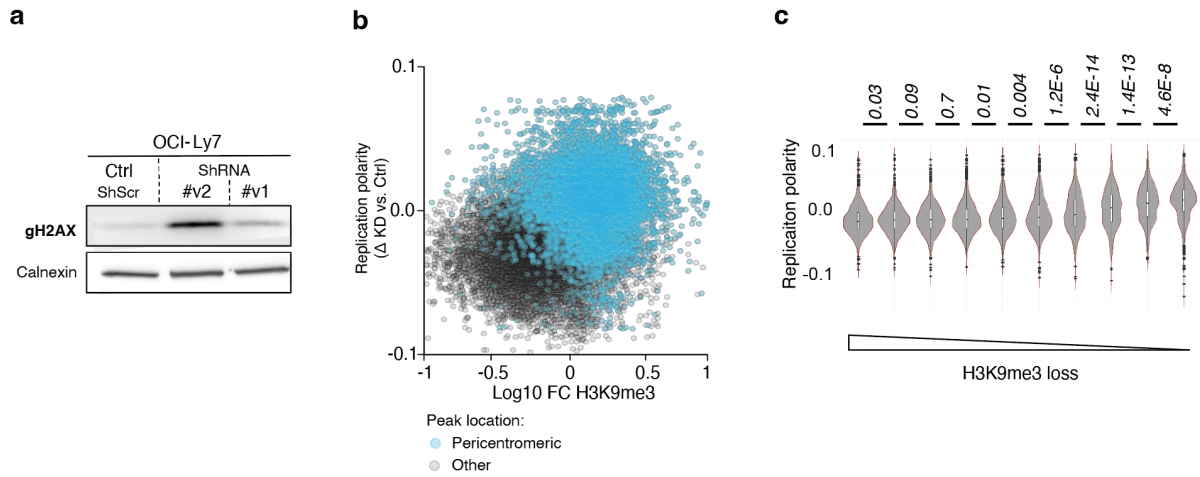


**Figure 10 | Figure S3 related to Figure 4: ZNF587/417 depletion alters the heterochromatin landscape of lymphoma cells.**

(a) Box plots showing the quartiles, the 10<sup>th</sup> and 90<sup>th</sup> percentiles (whiskers) of all, gaining, and losing H3k9me3 peaks lengths (in kilo-bases). Statistics: Two-sided Mann–Whitney U-test.

(b) Top: Venn diagrams of differentially expressed TFs (FDR < 0.05) upon ZNF587/417 KD in U2932 and OCI-Ly7 cells 6 days after LV transduction. The number of differentially expressed TFs shared between the two cell lines is shown in the overlapping area of the two diagrams. The number of TFs differentially expressed in only one of the two cell lines is shown in the non-overlapping area of the diagrams. Stacked bar charts depicting the proportion of up- and downregulated KZFPs amongst up- and downregulated TFs in U2932 and OCI-Ly7 are shown on each side of the Venn diagrams.

Bottom: Scatterplot of log2 fold changes of DE TFs shared between U2932 (x-axis) and OCI-Ly7 (y-axis) KD cells. Different colors were used to highlight TFs and KZFPs that were identified as positive hits from the Cox regression analysis screen reported in Figure 1.



**Figure 11 | Figure S4 related to Figure 5: ZNF587/417 depletion triggers replicative stress in lymphoma cells.**

(a) Western blot analysis of  $\gamma$ H2AX in OCI-Ly7 cells 3 days after LV transduction with 2 different anti-ZNF587/417 shRNAs and control shRNA (shScr). Calnexin was used as a loading control. (b) Scatter plot of replication polarity in function of H3K9me3 changes identical as Figure 5h. H3K9me3 peaks in pericentromeric regions are highlighted in blue and the ones in other regions in grey. (c) Violin plots of replication polarity deltas between KD and control conditions calculated as in Figure 5h grouped in 10 bins of the same numbers of H3K9me3 peaks and ranked by decreasing H3K9me3 loss. Statistics: Two-sided Mann–Whitney U-test.

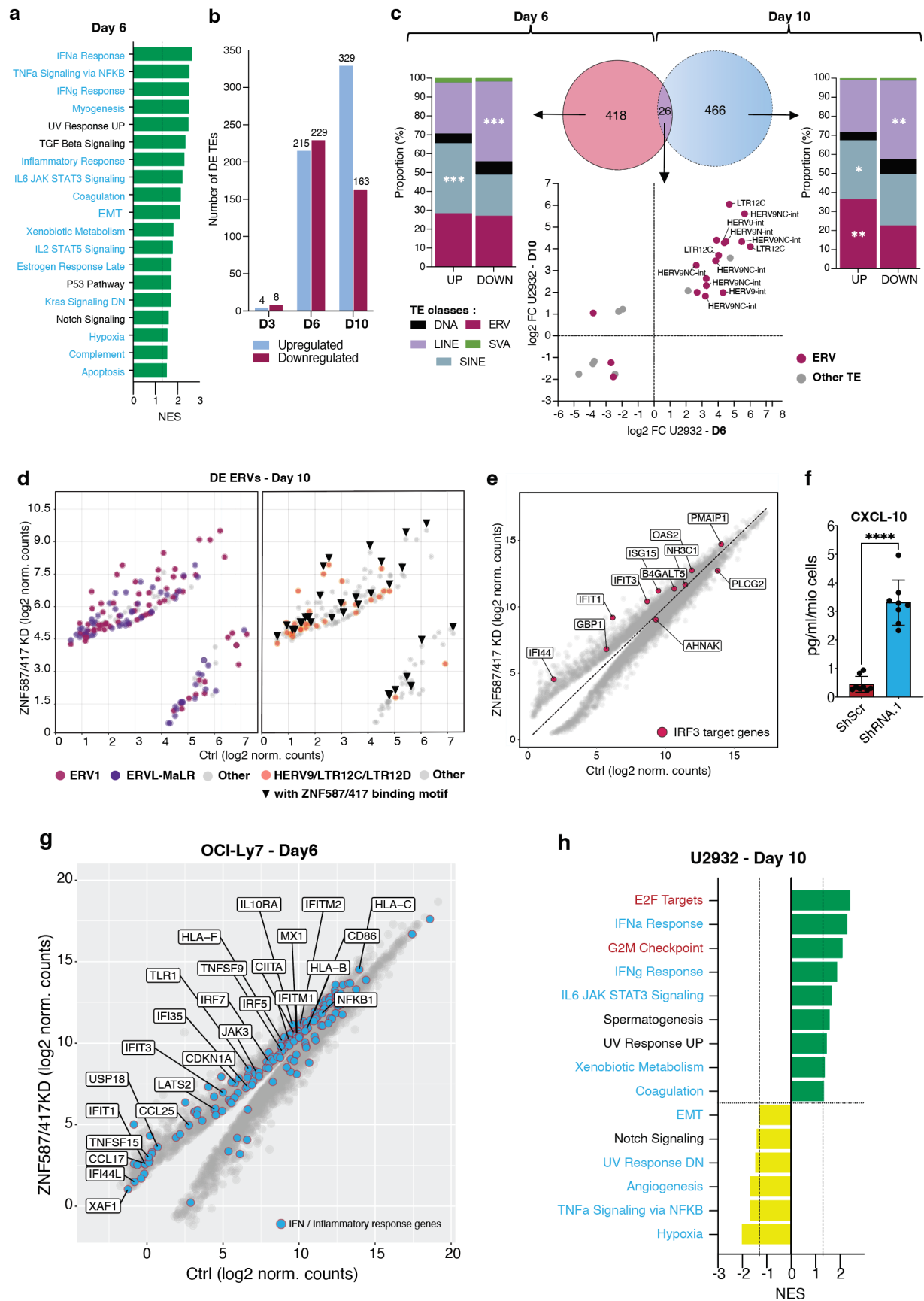


Figure 12 | **Figure S5 related to Figure 6: ZNF587/417 depletion leads to cell-intrinsic inflammation.**

(a) Waterfall plot of Hallmark GSEA signatures ranked by NES issued from RNA-seq data of day 6 ZNF587/417 KD in U2932 cells. Enriched signatures ( $NES > 0$ ) are highlighted in green. No signature was found significantly depleted. The dotted line represents p-value cutoff  $< 0.05$ . Hallmark labels found enriched in KZFP<sup>Low</sup> from Fig. 2 are highlighted in blue. (b) Bar plot depicting the number of DE TE integrants ( $FDR < 0.05$ ,  $FC > 2$ ) at day 3, 6, and 10 of ZNF587/417 KD in U2932 cells. Control shRNA-transduced cells collected in parallel to shRNA.1 KD cells were used as controls for every time point. (c) Center: Venn diagrams of DE TEs ( $FDR < 0.05$ ) upon ZNF587/417 KD in U2932 cells at day 6 (left) and 10 (right) after LV transduction. The number of DE TEs shared between the two time-points is shown at the intersection of the two disks. The number of TEs differentially expressed in only one of the two cell lines is shown in the non-overlapping area of the disks. Stacked bar charts depicting the proportion of up- and downregulated TEs at day 6 (left) and 10 (right) in U2932 KD cells are shown on each side of the Venn diagrams Bottom: Scatterplot of log<sub>2</sub> fold changes of DE TEs shared between day 6 (x-axis) and day 10 (y-axis) in U2932 KD cells (ERVs were highlighted in dark red and other TEs in gray dots). (d) Scatterplot of log<sub>2</sub> fold changes of DE ERVs at day 10 in U2932 KD cells. ERV1 elements are highlighted with dark red dots, ERVL-MaLR with purple dots, and other ERVs with gray dots. ERVs with a ZNF587 and/or ZNF417 binding motif are pointed with a triangle. The top 3 binding motifs determined in H1 embryonic stem cells were used<sup>154</sup> for both KZFPs. (e) Scatter plot of RNA-seq from ZNF417/587 U2932 KD versus (shScr) control cells at day 6 of KD, outlining DEGs (grey dots,  $FDR < 0.05$ ) and among them the genes reported as IRF3 targets by Grandvaux et al.<sup>264</sup> (dark red dots). (f) Bar plot measuring the concentration of CXCL10 cytokine by immunoassay in cell culture supernatant of ZNF587/417 shRNA.1 KD (blue) and control cells (dark red) 6 days after LV transduction normalized by the number of cells in each condition (pg/ml/millions of cells). (g) Scatter plot of RNA-seq from ZNF417/587 OCI-Ly7 KD versus control cells at day 6 of KD, outlining DEGs (grey dots,  $FDR < 0.05$ ) and among them genes belonging to type I/II IFN and Inflammatory response Hallmark gene sets, Interferon Signaling Reactome, and cellular response to type I, II and III IFN gene ontology terms (blue dots). (h) Waterfall plot of Hallmark GSEA signatures ranked by NES issued from RNA-seq data of day 10 ZNF587/417 KD in U2932 cells. Enriched signatures ( $NES > 0$ ) are highlighted in green and depleted signatures ( $NES < 0$ ) in yellow. The dotted line represents p-value cutoff  $< 0.05$ . Hallmark labels found enriched in KZFP-Low/-High groups from Fig. 2 are highlighted in blue and dark red, respectively.





# Perspectives

This study contributed to increase our understanding of KZFP function in DLBCL, by integrating transcriptomic and epigenetic data, pointing to the involvement of ZNF587/417 in DNA replication dynamics, genome stability, and cell-intrinsic inflammatory responses. It is the first study involving primate-specific KZFPs in the genome stability of cancer cells, thus encouraging further research regarding their upstream regulators and the implication of other KZFPs in immune evasion, cancer cell identity, and replicative stress regulation. The transcriptomic reprogramming of lymphoma cells surviving ZNF587/417 depletion, asks whether acute changes in their expression level may also occur during tumorigenesis and contribute to the emergence of more aggressive clones and/or immunity activation, through waves of endogenous RS. Experiments using humanized mice models may reveal if the inflammatory reaction triggered by their downregulation could be able to enhance anti-tumor immune responses or have the opposite effect. With the advent of Proteolysis targeting chimeric (PROTAC) therapies, the prospect of targeting KZFPs is now reachable, thus encouraging further research in this field.

# List of abbreviations

ABC	Activated B cell
ADCC	Antibody-Dependent Cellular Cytotoxicity
AID	Activation-Induced Deaminase
AKT	Serine/threonine kinase 1; protein kinase B
ASCT	Autologous Stem Cell Transplant
ATP	Adenosine Triphosphate
AV	Annexin V
BCL2	B-Cell Lymphoma 2
BCL6	B-Cell Lymphoma 6
BCR	B Cell Receptor
BCR-ABL	Breakpoint Cluster Region-Abelson Proto-oncogene
BL	Burkitt Lymphoma
bp	base pairs
C2H2	2 Cysteine – 2 Histidine
CAR	Chimeric Antigen Receptor
CD20	Cluster of Differentiation 20
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
cGAS	Cyclic GMP-AMP Synthase
CHM13	Complete Hydatidiform Mole 13
Chr	Chromosome
CIN	Chromosomal Instability
CML	Chronic Myelogenous Leukemia
CNA	Copy Number Alterations
COO	Cell Of Origin
CR	Complete Remission
CSR	Class-Switch Recombination
CUT&Tag	Cleavage Under Targets and Tagmentation
DEG	Differentially Expressed Gene
DLBCL	Diffuse Large B Cell Lymphoma
DNA	Deoxyribonucleic Acid
EdU	5-Ethynyl-2'-deoxyuridine
EMT	Epithelial-to-Mesenchymal Transition
ERFS	Early Replication Fragile Sites
ERV	Endogenous Retrovirus
FC	Fold Change
FDA	Food and Drug Administration
FDR	False Discovery Rate
GC	Germinal Center

GCB	Germinal Center B cell
gH2AX	gamma-H2A.X Variant Histon
GSEA	Gene Set Enrichment Analysis
H3K9me3	Histone 3 Lysine 9 methyl 3
HL	Hodgkin Lmyphoma
HLA	Human Leukocyte Antigen
HR	Hazard Ratio
IFN	Interferon
IgM	Immunoglobulin M
IRDS	Interferon-Related DNA Damage Signature
IRF	Interferon Regulatory Factor
KAP1	KRAB-Associated Protein 1 (a.k.a. TRIM28)
kb	kilo-base
KD	Knock Down
KRAB	Krüppel-Associated Box
KZFP	Krüppel-Associated Box Domain-Containing Zinc-Finger Protein
LDH	Lactate Dehydrogenase
LINE	Long Interspersed Elements
log10	logarithm at base 10
log2	logarithm at base 2
LTR	Long Terminal Repeats
mAbs	monoclonal antibodies
MFI	Median Fluorescence Intensity
MYC	Myelocytomatosis oncogene product
NES	Normalized Enrichment Score
NF-kB	Nuclear Factor-kappa B
NHL	Non-Hodgkin Lymphoma
NK	Natural Killer
PET-CT	Positron Emission Tomography – Computed Tomography
PI	Propidium Iodide
PI3K	Phosphatidylinositol-3 kinase
PMBCL	Primary Mediastinal B Cell Lymphoma
pol	polymerase
PR	Partial Remission
R-CHOP	Rituximab-Cyclophosphamide/Hydroxydaunorubicin/Oncovin/Prednisolone
R/R	Relapse/Refractory
REAL	Revised European-American Lymphoma
RNA	Ribonucleic Acid
RS	Replication Stress
SD	Standard Deviation
SEM	Standard Error of the Mean
SETDB1	Su(var)3-9 Enhancer-of-zeste and Trithorax Domain Bifurcated Histone Lysine Methyltransferase 1

SHM	Somatic Hypermutation
shRNA	small hairpin Ribonucleic Acid
shScr	small hairpin Scramble
SINE	Short Interspersed Elements
STING	Stimulator of Interferon Genes
SVA	SINE-VNTR-Alu
T2T	Telomere-to-Telomere
TCGA	The Cancer Genome Atlas
TCR	T Cell Receptor
TE	Transposable Element
TEeRS	TE embedded Regulatory Sequences
TF	Transcription factor
Tfh	Follicular Helper T cells
TNF $\alpha$	Tumor Necrosis Factor alpha
TP53	Tumor Protein P53
TrAEL-Seq	Transferase-Activated End Ligation sequencing
TRIM28	Tripartite motif-containing 28 (a.k.a. KAP1)
US	United States
VNTR	Variable Number of Tandem Repeats
WES	Whole-Exome Sequencing
WHO	World Health Organization
ZNF	Zinc finger

# REFERENCES

(The references for this paper have been merged with the Bibliography of the thesis.)

1. Hodgkin Thomas. On some Morbid Appearances of the Absorbent Glands and Spleen. *Med Chir Trans.* 1832;17:68-114. doi:10.1177/095952873201700106
2. Billroth Theod. Multiple lymphome. Erfolgreiche behandlung mit arsenik. *Wien Med Wochenschr.* 1871;21:1066.
3. Sternberg C. Über eine eigenartige unter dem Bilde der Pseudoleukämie verlaufende Tuberculose des lymphatischen Apparates. *Ztschr Heilk.* 1898;19:21-90.
4. Reed DM. On the pathological changes in Hodgkin's disease, with special reference to its relation to tuberculosis. *John Hopkins Hosp Rep.* 1902;10:133-196.
5. Rappaport H. Washington DC, USA: United States Armed Forces Institute of Pathology. Tumors of the haematopoeitic system: Section 3 3, fascicle 8; p. 442. In: *Atlas of Tumor Pathology.* ; 1966.
6. Thandra KC, Barsouk A, Saginala K, Padala SA, Barsouk A, Rawla P. Epidemiology of Non-Hodgkin's Lymphoma. *Med Sci (Basel).* 2021;9(1):5. doi:10.3390/medsci9010005
7. Lukes RJ, Collins RD. Immunologic characterization of human malignant lymphomas. *Cancer.* 1974;34(4 Suppl):suppl:1488-1503. doi:10.1002/1097-0142(197410)34:8+<1488::aid-cnrcr2820340822>3.0.co;2-c
8. Harris NL, Jaffe ES, Stein H, et al. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood.* 1994;84(5):1361-1392.
9. Xie Y, Pittaluga S, Jaffe ES. The histological classification of diffuse large B-cell lymphomas. *Semin Hematol.* 2015;52(2):57-66. doi:10.1053/j.seminhematol.2015.01.006
10. Rosenwald A, Staudt LM. Gene Expression Profiling of Diffuse Large B-Cell Lymphoma. *Leukemia & Lymphoma.* 2003;44(sup3):S41-S47. doi:10.1080/10428190310001623775
11. Lenz G, Wright GW, Emre NCT, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci USA.* 2008;105(36):13520-13525. doi:10.1073/pnas.0804295105
12. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature.* 2001;412(6844):341-346. doi:10.1038/35085588
13. Küppers R, Rajewsky K, Hansmann ML. Diffuse large cell lymphomas are derived from mature B cells carrying V region genes with a high load of somatic mutation and evidence of selection for antibody expression. *Eur J Immunol.* 1997;27(6):1398-1405. doi:10.1002/eji.1830270616

14. Lenz G, Nagel I, Siebert R, et al. Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *Journal of Experimental Medicine*. 2007;204(3):633-643. doi:10.1084/jem.20062041
15. Morin RD, Johnson NA, Severson TM, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet*. 2010;42(2):181-185. doi:10.1038/ng.518
16. Muppidi JR, Schmitz R, Green JA, et al. Loss of signalling via Gα13 in germinal centre B-cell-derived lymphoma. *Nature*. 2014;516(7530):254-258. doi:10.1038/nature13765
17. Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*. 2009;459(7247):717-721. doi:10.1038/nature07968
18. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*. 2022;36(7):1720-1748. doi:10.1038/s41375-022-01620-2
19. El-Galaly TC, Villa D, Gormsen LC, Baech J, Lo A, Cheah CY. FDG-PET/CT in the management of lymphomas: current status and future directions. *J Intern Med*. 2018;284(4):358-376. doi:10.1111/joim.12813
20. Gilbert R. NOTRE MÉTHODE DE ROENTGENTHÉRAPIE DE LA LYMPHOGRANULOMATOSE (HODGKIN): RÉSULTATS ÉLOIGNÉS. *Acta Radiologica*. 1931;12(6).
21. Gilman A, Philips FS. The Biological Actions and Therapeutic Applications of the B-Chloroethyl Amines and Sulfides. *Science*. 1946;103(2675):409-436. doi:10.1126/science.103.2675.409
22. Fisher RI, Gaynor ER, Dahlberg S, et al. Comparison of a Standard Regimen (CHOP) with Three Intensive Chemotherapy Regimens for Advanced Non-Hodgkin's Lymphoma. *N Engl J Med*. 1993;328(14):1002-1006. doi:10.1056/NEJM199304083281404
23. Di Marco A, Gaetani M, Orezzi P, et al. 'Daunomycin', a New Antibiotic of the Rhodomycin Group. *Nature*. 1964;201(4920):706-707. doi:10.1038/201706a0
24. Pommier Y, Leo E, Zhang H, Marchand C. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chemistry & Biology*. 2010;17(5):421-433. doi:10.1016/j.chembiol.2010.04.012
25. Miller TP, Dahlberg S. Treatment of diffuse large-cell lymphoma: a summary of outcome for patients treated by Southwest Oncology Group. *Cancer Treat Res*. 1993;66:53-63. doi:10.1007/978-1-4615-3084-8\_4
26. Milstein C. The Wellcome Foundation Lecture, 1980: Monoclonal Antibodies from Hybrid Myelomas. *Proceedings of the Royal Society of London Series B, Biological Sciences*. 1981;211(1185):393-412.

27. Nadler LM, Stashenko P, Hardy R, et al. Serotherapy of a patient with a monoclonal antibody directed against a human lymphoma-associated antigen. *Cancer Res.* 1980;40(9):3147-3154.
28. Nadler LM, Ritz J, Hardy R, Pesando JM, Schlossman SF, Stashenko P. A unique cell surface antigen identifying lymphoid malignancies of B cell origin. *J Clin Invest.* 1981;67(1):134-140. doi:10.1172/JCI110005
29. Tedder TF, Streuli M, Schlossman SF, Saito H. Isolation and structure of a cDNA encoding the B1 (CD20) cell-surface antigen of human B lymphocytes. *Proc Natl Acad Sci U S A.* 1988;85(1):208-212. doi:10.1073/pnas.85.1.208
30. Reff M, Carner K, Chambers K, et al. Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. *Blood.* 1994;83(2):435-445. doi:10.1182/blood.V83.2.435.435
31. Maloney DG, Grillo-López AJ, White CA, et al. IDEC-C2B8 (Rituximab) Anti-CD20 Monoclonal Antibody Therapy in Patients With Relapsed Low-Grade Non-Hodgkin's Lymphoma. *Blood.* 1997;90(6):2188-2195. doi:10.1182/blood.V90.6.2188
32. McLaughlin P, Grillo-López AJ, Link BK, et al. Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program. *J Clin Oncol.* 1998;16(8):2825-2833. doi:10.1200/JCO.1998.16.8.2825
33. James JS, Dubs G. FDA approves new kind of lymphoma treatment. Food and Drug Administration. *AIDS Treat News.* 1997;(No 284):2-3.
34. Coiffier B, Lepage E, Brière J, et al. CHOP Chemotherapy plus Rituximab Compared with CHOP Alone in Elderly Patients with Diffuse Large-B-Cell Lymphoma. *N Engl J Med.* 2002;346(4):235-242. doi:10.1056/NEJMoa011795
35. Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood.* 2010;116(12):2040-2045. doi:10.1182/blood-2010-03-276246
36. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329(14):987-994. doi:10.1056/NEJM199309303291402
37. Sehn LH, Berry B, Chhanabhai M, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood.* 2007;109(5):1857-1861. doi:10.1182/blood-2006-08-038257
38. Burggraaff CN, de Jong A, Hoekstra OS, et al. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2019;46(1):65-79. doi:10.1007/s00259-018-4103-3



39. Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. *N Engl J Med.* 2001;344(14):1031-1037. doi:10.1056/NEJM200104053441401
40. Honigberg LA, Smith AM, Sirisawad M, et al. The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc Natl Acad Sci U S A.* 2010;107(29):13075-13080. doi:10.1073/pnas.1004594107
41. Salles G, Schuster SJ, de Vos S, et al. Efficacy and safety of idelalisib in patients with relapsed, rituximab- and alkylating agent-refractory follicular lymphoma: a subgroup analysis of a phase 2 study. *Haematologica.* 2017;102(4):e156-e159. doi:10.3324/haematol.2016.151738
42. Eyre TA, Walter HS, Iyengar S, et al. Efficacy of venetoclax monotherapy in patients with relapsed, refractory mantle cell lymphoma after Bruton tyrosine kinase inhibitor therapy. *Haematologica.* 2019;104(2):e68-e71. doi:10.3324/haematol.2018.198812
43. Kale J, Osterlund EJ, Andrews DW. BCL-2 family proteins: changing partners in the dance towards death. *Cell Death Differ.* 2018;25(1):65-80. doi:10.1038/cdd.2017.186
44. Gross G, Gorochov G, Waks T, Eshhar Z. Generation of effector T cells expressing chimeric T cell receptor with antibody type-specificity. *Transplant Proc.* 1989;21(1 Pt 1):127-130.
45. Gross G, Waks T, Eshhar Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. *Proc Natl Acad Sci U S A.* 1989;86(24):10024-10028. doi:10.1073/pnas.86.24.10024
46. Eshhar Z, Waks T, Gross G, Schindler DG. Specific activation and targeting of cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains and the gamma or zeta subunits of the immunoglobulin and T-cell receptors. *Proc Natl Acad Sci U S A.* 1993;90(2):720-724. doi:10.1073/pnas.90.2.720
47. Grupp SA, Kalos M, Barrett D, et al. Chimeric Antigen Receptor–Modified T Cells for Acute Lymphoid Leukemia. *N Engl J Med.* 2013;368(16):1509-1518. doi:10.1056/NEJMoa1215134
48. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med.* 2017;377(26):2531-2544. doi:10.1056/NEJMoa1707447
49. Schuster SJ, Bishop MR, Tam CS, et al. Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N Engl J Med.* 2019;380(1):45-56. doi:10.1056/NEJMoa1804980
50. Tan Su Yin E, Hu YX, Huang H. The breakthrough and the future: CD20 chimeric antigen receptor T-cell therapy for hematologic malignancies. *ImmunoMedicine.* 2022;2(1). doi:10.1002/imed.1039
51. Kanas G, Ge W, Quek RGW, Keeven K, Nersesyan K, Jon E. Arnason. Epidemiology of diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) in the United

- States and Western Europe: population-level projections for 2020–2025. *Leukemia & Lymphoma*. 2022;63(1):54-63. doi:10.1080/10428194.2021.1975188
52. Méchali M. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol*. 2010;11(10):728-738. doi:10.1038/nrm2976
  53. Yeeles JTP, Deegan TD, Janska A, Early A, Diffley JFX. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*. 2015;519(7544):431-435. doi:10.1038/nature14285
  54. Aparicio T, Guillou E, Coloma J, Montoya G, Méndez J. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res*. 2009;37(7):2087-2095. doi:10.1093/nar/gkp065
  55. Garg P, Burgers PMJ. DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol*. 2005;40(2):115-128. doi:10.1080/10409230590935433
  56. Johnson A, O'Donnell M. Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem*. 2005;74:283-315. doi:10.1146/annurev.biochem.73.011303.073859
  57. Boos D, Ferreira P. Origin Firing Regulations to Control Genome Replication Timing. *Genes (Basel)*. 2019;10(3):199. doi:10.3390/genes10030199
  58. Petropoulos M, Champeris Tsaniras S, Taraviras S, Lygerou Z. Replication Licensing Aberrations, Replication Stress, and Genomic Instability. *Trends in Biochemical Sciences*. 2019;44(9):752-764. doi:10.1016/j.tibs.2019.03.011
  59. Fragkos M, Ganier O, Coulombe P, Méchali M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol*. 2015;16(6):360-374. doi:10.1038/nrm4002
  60. Petryk N, Kahli M, d'Aubenton-Carafa Y, et al. Replication landscape of the human genome. *Nat Commun*. 2016;7:10208. doi:10.1038/ncomms10208
  61. Zhang J, MacLennan ICM, Liu YJ, Lane PJJ. Is rapid proliferation in B centroblasts linked to somatic mutation in memory B cell clones? *Immunology Letters*. 1988;18(4):297-299. doi:10.1016/0165-2478(88)90178-2
  62. Gitlin AD, Mayer CT, Oliveira TY, et al. HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells. *Science*. 2015;349(6248):643-646. doi:10.1126/science.aac4919
  63. Hübschmann D, Kleinheinz K, Wagener R, et al. Mutational mechanisms shaping the coding and noncoding genome of germinal center derived B-cell lymphomas. *Leukemia*. 2021;35(7):2002-2016. doi:10.1038/s41375-021-01251-z
  64. Murga M, Campaner S, Lopez-Contreras AJ, et al. Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nat Struct Mol Biol*. 2011;18(12):1331-1335. doi:10.1038/nsmb.2189

65. Zeman MK, Cimprich KA. Causes and consequences of replication stress. *Nat Cell Biol.* 2014;16(1):2-9. doi:10.1038/ncb2897
66. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science.* 2003;300(5625):1542-1548. doi:10.1126/science.1083430
67. Vassin VM, Anantha RW, Sokolova E, Kanner S, Borowiec JA. Human RPA phosphorylation by ATR stimulates DNA synthesis and prevents ssDNA accumulation during DNA-replication stress. *J Cell Sci.* 2009;122(Pt 22):4070-4080. doi:10.1242/jcs.053702
68. Toledo L, Neelsen KJ, Lukas J. Replication Catastrophe: When a Checkpoint Fails because of Exhaustion. *Molecular Cell.* 2017;66(6):735-749. doi:10.1016/j.molcel.2017.05.001
69. Kitao H, Iimori M, Kataoka Y, et al. DNA replication stress and cancer chemotherapy. *Cancer Sci.* 2018;109(2):264-271. doi:10.1111/cas.13455
70. Marinello J, Delcuratolo M, Capranico G. Anthracyclines as Topoisomerase II Poisons: From Early Studies to New Perspectives. *IJMS.* 2018;19(11):3480. doi:10.3390/ijms19113480
71. Collins A, Oates DJ. Hydroxyurea: effects on deoxyribonucleotide pool sizes correlated with effects on DNA repair in mammalian cells. *Eur J Biochem.* 1987;169(2):299-305. doi:10.1111/j.1432-1033.1987.tb13612.x
72. Halazonetis TD, Gorgoulis VG, Bartek J. An Oncogene-Induced DNA Damage Model for Cancer Development. *Science.* 2008;319(5868):1352-1355. doi:10.1126/science.1140735
73. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. Human cancers express a mutator phenotype. *Proc Natl Acad Sci U S A.* 2006;103(48):18238-18242. doi:10.1073/pnas.0607057103
74. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer.* 2011;11(6):450-457. doi:10.1038/nrc3063
75. Knijnenburg TA, Wang L, Zimmermann MT, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Reports.* 2018;23(1):239-254.e6. doi:10.1016/j.celrep.2018.03.076
76. Nigg EA. Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays.* 1995;17(6):471-480. doi:10.1002/bies.950170603
77. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell.* 1997;88(5):593-602. doi:10.1016/S0092-8674(00)81902-9
78. Jones RM, Mortusewicz O, Afzal I, et al. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene.* 2013;32(32):3744-3753. doi:10.1038/onc.2012.387

79. Frum RA, Singh S, Vaughan C, et al. The human oncoprotein MDM2 induces replication stress eliciting early intra-S-phase checkpoint response and inhibition of DNA replication origin firing. *Nucleic Acids Res.* 2014;42(2):926-940. doi:10.1093/nar/gkt944
80. Dominguez-Sola D, Ying CY, Grandori C, et al. Non-transcriptional control of DNA replication by c-Myc. *Nature.* 2007;448(7152):445-451. doi:10.1038/nature05953
81. Bester AC, Roniger M, Oren YS, et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell.* 2011;145(3):435-446. doi:10.1016/j.cell.2011.03.044
82. Ekholm-Reed S, Méndez J, Tedesco D, Zetterberg A, Stillman B, Reed SI. Deregulation of cyclin E in human cells interferes with prereplication complex assembly. *J Cell Biol.* 2004;165(6):789-800. doi:10.1083/jcb.200404092
83. Beck H, Nähse-Kumpf V, Larsen MSY, et al. Cyclin-dependent kinase suppression by WEE1 kinase protects the genome through control of replication initiation and nucleotide consumption. *Mol Cell Biol.* 2012;32(20):4226-4236. doi:10.1128/MCB.00412-12
84. Elbæk CR, Petrosius V, Benada J, Erichsen L, Damgaard RB, Sørensen CS. WEE1 kinase protects the stability of stalled DNA replication forks by limiting CDK2 activity. *Cell Reports.* 2022;38(3):110261. doi:10.1016/j.celrep.2021.110261
85. Sutherland GR. Heritable fragile sites on human chromosomes I. Factors affecting expression in lymphocyte culture. *Am J Hum Genet.* 1979;31(2):125-135.
86. Arlt MF, Casper AM, Glover TW. Common fragile sites. *Cytogenet Genome Res.* 2003;100(1-4):92-100. doi:10.1159/000072843
87. Chen YH, Keegan S, Kahli M, et al. Transcription shapes DNA replication initiation and termination in human cells. *Nat Struct Mol Biol.* 2019;26(1):67-77. doi:10.1038/s41594-018-0171-0
88. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet.* 2015;16(10):583-597. doi:10.1038/nrg3961
89. Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet.* 2016;48(11):1385-1395. doi:10.1038/ng.3672
90. Manzo SG, Hartono SR, Sanz LA, et al. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* 2018;19(1):100. doi:10.1186/s13059-018-1478-1
91. Barroso S, Herrera-Moyano E, Muñoz S, García-Rubio M, Gómez-González B, Aguilera A. The DNA damage response acts as a safeguard against harmful DNA-RNA hybrids of different origins. *EMBO Rep.* 2019;20(9):e47250. doi:10.15252/embr.201847250

92. Bayona-Feliu A, Barroso S, Muñoz S, Aguilera A. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription–replication conflicts. *Nat Genet.* 2021;53(7):1050-1063. doi:10.1038/s41588-021-00867-2
93. De Magis A, Manzo SG, Russo M, et al. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc Natl Acad Sci USA.* 2019;116(3):816-825. doi:10.1073/pnas.1810409116
94. Šviković S, Crisp A, Tan-Wong SM, et al. R-loop formation during S phase is restricted by PrimPol-mediated repriming. *EMBO J.* 2019;38(3). doi:10.15252/embj.201899793
95. Feng Y, Seija N, Di Noia JM, Martin A. AID in Antibody Diversification: There and Back Again. *Trends in Immunology.* 2020;41(7):586-600. doi:10.1016/j.it.2020.04.009
96. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol.* 2003;4(5):442-451. doi:10.1038/ni919
97. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol.* 2004;172(6):3382-3384. doi:10.4049/jimmunol.172.6.3382
98. Pettersen HS, Galashevskaya A, Doseth B, et al. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair (Amst).* 2015;25:60-71. doi:10.1016/j.dnarep.2014.11.006
99. Liu YJ, Zhang J, Lane PJ, Chan EY, MacLennan IC. Sites of specific B cell activation in primary and secondary responses to T cell-dependent and T cell-independent antigens. *Eur J Immunol.* 1991;21(12):2951-2962. doi:10.1002/eji.1830211209
100. Xia Y, Zhang X. The Spectrum of MYC Alterations in Diffuse Large B-Cell Lymphoma. *Acta Haematol.* 2020;143(6):520-528. doi:10.1159/000505892
101. Karube K, Campo E. MYC Alterations in Diffuse Large B-Cell Lymphomas. *Seminars in Hematology.* 2015;52(2):97-106. doi:10.1053/j.seminhematol.2015.01.009
102. Johnson NA, Slack GW, Savage KJ, et al. Concurrent Expression of MYC and BCL2 in Diffuse Large B-Cell Lymphoma Treated With Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone. *JCO.* 2012;30(28):3452-3459. doi:10.1200/JCO.2011.41.0985
103. Baluapuri A, Hofstetter J, Dudvarski Stankovic N, et al. MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Molecular Cell.* 2019;74(4):674-687.e11. doi:10.1016/j.molcel.2019.02.031
104. Barlow JH, Faryabi RB, Callén E, et al. Identification of early replicating fragile sites that contribute to genome instability. *Cell.* 2013;152(3):620-632. doi:10.1016/j.cell.2013.01.006

105. Srinivasan SV, Dominguez-Sola D, Wang LC, Hyrien O, Gautier J. Cdc45 is a critical effector of myc-dependent DNA replication stress. *Cell Rep.* 2013;3(5):1629-1639. doi:10.1016/j.celrep.2013.04.002
106. Restelli V, Vagni M, Arribas AJ, Bertoni F, Damia G, Carrassa L. Inhibition of CHK1 and WEE1 as a new therapeutic approach in diffuse large B cell lymphomas with MYC deregulation. *Br J Haematol.* 2018;181(1):129-133. doi:10.1111/bjh.14506
107. Young LA, O'Connor LO, de Renty C, et al. Differential Activity of ATR and WEE1 Inhibitors in a Highly Sensitive Subpopulation of DLBCL Linked to Replication Stress. *Cancer Research.* 2019;79(14):3762-3775. doi:10.1158/0008-5472.CAN-18-2480
108. Guilbaud G, Murat P, Wilkes HS, Lerner LK, Sale JE, Krude T. Determination of human DNA replication origin position and efficiency reveals principles of initiation zone organisation. *Nucleic Acids Research.* 2022;50(13):7436-7450. doi:10.1093/nar/gkac555
109. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell.* 2007;128(4):735-745. doi:10.1016/j.cell.2007.02.009
110. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol.* 2018;19(4):229-244. doi:10.1038/nrm.2017.119
111. Ninova M, Godneeva B, Chen YCA, et al. The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. *Molecular Cell.* 2020;77(3):571-585.e4. doi:10.1016/j.molcel.2019.09.033
112. Leclerc S, Kitagawa K. The Role of Human Centromeric RNA in Chromosome Stability. *Front Mol Biosci.* 2021;8:642732. doi:10.3389/fmolb.2021.642732
113. Zagris N, Podimatas T. 5-Azacytidine changes gene expression and causes developmental arrest of early chick embryo. *Int J Dev Biol.* 1994;38(4):741-744.
114. Taylor SM. 5-Aza-2'-deoxycytidine: cell differentiation and DNA methylation. *Leukemia.* 1993;7 Suppl 1:3-8.
115. Fandy TE, Jiemjit A, Thakar M, Rhoden P, Suarez L, Gore SD. Decitabine Induces Delayed Reactive Oxygen Species (ROS) Accumulation in Leukemia Cells and Induces the Expression of ROS Generating Enzymes. *Clinical Cancer Research.* 2014;20(5):1249-1258. doi:10.1158/1078-0432.CCR-13-1453
116. Hajek M, Biktasova A, Sewell A, et al. Global Genome Demethylation Causes Transcription-Associated DNA Double Strand Breaks in HPV-Associated Head and Neck Cancer Cells. *Cancers (Basel).* 2020;13(1):21. doi:10.3390/cancers13010021
117. Montavon T, Shukeir N, Erikson G, et al. Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun.* 2021;12(1):4359. doi:10.1038/s41467-021-24532-8

118. Južnić L, Peuker K, Strigli A, et al. SETDB1 is required for intestinal epithelial differentiation and the prevention of intestinal inflammation. *Gut*. 2021;70(3):485-498. doi:10.1136/gutjnl-2020-321339
119. Gao C, Wang X, Chen L, Wang JH, Gao ZT, Wang H. Knockdown of Bcl-3 inhibits cell growth and induces DNA damage in HTLV-1-infected cells. *Asian Pac J Cancer Prev*. 2013;14(1):405-408. doi:10.7314/apjcp.2013.14.1.405
120. Parker C, Chambers AC, Flanagan DJ, et al. BCL-3 loss sensitises colorectal cancer cells to DNA damage by targeting homologous recombination. *DNA Repair*. 2022;115:103331. doi:10.1016/j.dnarep.2022.103331
121. Härtlova A, Erttmann SF, Raffi FAM, et al. DNA Damage Primes the Type I Interferon System via the Cytosolic DNA Sensor STING to Promote Anti-Microbial Innate Immunity. *Immunity*. 2015;42(2):332-343. doi:10.1016/j.immuni.2015.01.012
122. Erdal E, Haider S, Rehwinkel J, Harris AL, McHugh PJ. A prosurvival DNA damage-induced cytoplasmic interferon response is mediated by end resection factors and is limited by Trex1. *Genes Dev*. 2017;31(4):353-369. doi:10.1101/gad.289769.116
123. Jazayeri A, Balestrini A, Garner E, Haber JE, Costanzo V. Mre11–Rad50–Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J*. 2008;27(14):1953-1962. doi:10.1038/emboj.2008.128
124. Wolf C, Rapp A, Berndt N, et al. RPA and Rad51 constitute a cell intrinsic mechanism to protect the cytosol from self DNA. *Nat Commun*. 2016;7:11752. doi:10.1038/ncomms11752
125. Bakhoun SF, Landau DA. Chromosomal Instability as a Driver of Tumor Heterogeneity and Evolution. *Cold Spring Harb Perspect Med*. 2017;7(6):a029611. doi:10.1101/cshperspect.a029611
126. Mackenzie KJ, Carroll P, Martin CA, et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature*. 2017;548(7668):461-465. doi:10.1038/nature23449
127. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science*. 2013;339(6121):786-791. doi:10.1126/science.1232458
128. Abe T, Barber GN. Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF- $\kappa$ B activation through TBK1. *J Virol*. 2014;88(10):5328-5341. doi:10.1128/JVI.00037-14
129. Coquel F, Neumayer C, Lin YL, Pasero P. SAMHD1 and the innate immune response to cytosolic DNA during DNA replication. *Current Opinion in Immunology*. 2019;56:24-30. doi:10.1016/j.coi.2018.09.017
130. Aicardi J, Goutières F. A Progressive familial encephalopathy in infancy with calcifications of the basal ganglia and chronic cerebrospinal fluid lymphocytosis: Aicardi and Goutieres: Basal Ganglia Calcifications. *Ann Neurol*. 1984;15(1):49-54. doi:10.1002/ana.410150109

131. Crow YJ, Chase DS, Lowenstein Schmidt J, et al. Characterization of human disease phenotypes associated with mutations in *TREX1* , *RNASEH2A* , *RNASEH2B* , *RNASEH2C* , *SAMHD1* , *ADAR* , and *IFIH1*. *Am J Med Genet.* 2015;167(2):296-312. doi:10.1002/ajmg.a.36887
132. Yang YG, Lindahl T, Barnes DE. Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. *Cell.* 2007;131(5):873-886. doi:10.1016/j.cell.2007.10.017
133. Vanpouille-Box C, Alard A, Aryankalayil MJ, et al. DNA exonuclease Trex1 regulates radiotherapy-induced tumour immunogenicity. *Nat Commun.* 2017;8:15618. doi:10.1038/ncomms15618
134. Vanpouille-Box C, Formenti SC, Demaria S. TREX1 dictates the immune fate of irradiated cancer cells. *OncolImmunology.* 2017;6(9):e1339857. doi:10.1080/2162402X.2017.1339857
135. Schott K, Majer C, Bulashevskaya A, et al. SAMHD1 in cancer: curse or cure? *J Mol Med (Berl).* 2022;100(3):351-372. doi:10.1007/s00109-021-02131-w
136. Prati B, da Silva Abjaude W, Termini L, et al. Three Prime Repair Exonuclease 1 (TREX1) expression correlates with cervical cancer cells growth in vitro and disease progression in vivo. *Sci Rep.* 2019;9(1):351. doi:10.1038/s41598-018-37064-x
137. Ho SSW, Zhang WYL, Tan NYJ, et al. The DNA Structure-Specific Endonuclease MUS81 Mediates DNA Sensor STING-Dependent Host Rejection of Prostate Cancer Cells. *Immunity.* 2016;44(5):1177-1189. doi:10.1016/j.immuni.2016.04.010
138. Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. *Cell.* 2018;172(4):650-665. doi:10.1016/j.cell.2018.01.029
139. Wang Z, Wang P, Li Y, et al. Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. *Sig Transduct Target Ther.* 2021;6(1):24. doi:10.1038/s41392-020-00422-1
140. Solimini NL, Luo J, Elledge SJ. Non-Oncogene Addiction and the Stress Phenotype of Cancer Cells. *Cell.* 2007;130(6):986-988. doi:10.1016/j.cell.2007.09.007
141. Mantovani F, Collavin L, Del Sal G. Mutant p53 as a guardian of the cancer cell. *Cell Death Differ.* 2019;26(2):199-212. doi:10.1038/s41418-018-0246-9
142. Willis A, Jung EJ, Wakefield T, Chen X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene.* 2004;23(13):2330-2338. doi:10.1038/sj.onc.1207396
143. Vogel MJ, Guelen L, de Wit E, et al. Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* 2006;16(12):1493-1504. doi:10.1101/gr.5391806
144. Huntley S, Baggott DM, Hamilton AT, et al. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large



- family of transcriptional repressors. *Genome Res.* 2006;16(5):669-677. doi:10.1101/gr.4842106
145. Peng H, Begg GE, Schultz DC, et al. Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol.* 2000;295(5):1139-1162. doi:10.1006/jmbi.1999.3402
  146. Friedman JR, Fredericks WJ, Jensen DE, et al. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* 1996;10(16):2067-2078. doi:10.1101/gad.10.16.2067
  147. Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res.* 1996;24(24):4859-4867. doi:10.1093/nar/24.24.4859
  148. Santoni de Sio FR, Massacand J, Barde I, et al. KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood.* 2012;119(20):4675-4685. doi:10.1182/blood-2011-12-401117
  149. Chen W, Schwalie PC, Pankevich EV, et al. ZFP30 promotes adipogenesis through the KAP1-mediated activation of a retrotransposon-derived Pparg2 enhancer. *Nat Commun.* 2019;10(1):1809. doi:10.1038/s41467-019-09803-9
  150. Helleboid PY, Heusel M, Duc J, et al. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. *EMBO J.* 2019;38(18):e101220. doi:10.15252/emboj.2018101220
  151. Zorzan I, Pellegrini M, Arboit M, et al. The transcriptional regulator ZNF398 mediates pluripotency and epithelial character downstream of TGF-beta in human PSCs. *Nat Commun.* 2020;11(1):2364. doi:10.1038/s41467-020-16205-9
  152. Pontis J, Planet E, Offner S, et al. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell.* 2019;24(5):724-735.e5. doi:10.1016/j.stem.2019.03.012
  153. Pontis J, Pulver C, Playfoot CJ, et al. Primate-specific transposable elements shape transcriptional networks during human development. *Nat Commun.* 2022;13(1):7178. doi:10.1038/s41467-022-34800-w
  154. Turelli P, Playfoot C, Grun D, et al. Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons. *Sci Adv.* 2020;6(35):eaba3200. doi:10.1126/sciadv.aba3200
  155. Hancks DC, Kazazian HH. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev.* 2012;22(3):191-203. doi:10.1016/j.gde.2012.02.006
  156. Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature.* 2017;543(7646):550-554. doi:10.1038/nature21683

157. Iouranova A, Grun D, Rossy T, et al. KRAB zinc finger protein ZNF676 controls the transcriptional influence of LTR12-related endogenous retrovirus sequences. *Mobile DNA*. 2022;13(1):4. doi:10.1186/s13100-021-00260-0
158. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. *Virology*. 2011;411(2):273-287. doi:10.1016/j.virol.2010.12.007
159. Ecco G, Cassano M, Kauzlaric A, et al. Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev Cell*. 2016;36(6):611-623. doi:10.1016/j.devcel.2016.02.024
160. Ito J, Kimura I, Soper A, et al. *Endogenous Retroviruses Drive KRAB Zinc-Finger Family Protein Expression for Tumor Suppression*. Cancer Biology; 2020. doi:10.1101/2020.02.02.931501
161. Huang C, Jia Y, Yang S, et al. Characterization of ZNF23, a KRAB-containing protein that is downregulated in human cancers and inhibits cell cycle progression. *Experimental Cell Research*. 2007;313(2):254-263. doi:10.1016/j.yexcr.2006.10.009
162. Cheng Y, Geng H, Cheng SH, et al. KRAB Zinc Finger Protein ZNF382 Is a Proapoptotic Tumor Suppressor That Represses Multiple Oncogenes and Is Commonly Silenced in Multiple Carcinomas. *Cancer Research*. 2010;70(16):6516-6526. doi:10.1158/0008-5472.CAN-09-4566
163. Cheng Y, Liang P, Geng H, et al. A Novel 19q13 Nucleolar Zinc Finger Protein Suppresses Tumor Cell Growth through Inhibiting Ribosome Biogenesis and Inducing Apoptosis but Is Frequently Silenced in Multiple Carcinomas. *Molecular Cancer Research*. 2012;10(7):925-936. doi:10.1158/1541-7786.MCR-11-0594
164. Li J, Hao N, Han J, Zhang M, Li X, Yang N. ZKSCAN3 drives tumor metastasis via integrin  $\beta$ 4/FAK/AKT mediated epithelial-mesenchymal transition in hepatocellular carcinoma. *Cancer Cell Int*. 2020;20:216. doi:10.1186/s12935-020-01307-7
165. Chauvin A, Bergeron D, Vencic J, et al. Downregulation of KRAB zinc finger proteins in 5-fluorouracil resistant colorectal cancer cells. *BMC Cancer*. 2022;22(1):363. doi:10.1186/s12885-022-09417-3
166. Sobocińska J, Molenda S, Machnik M, Oleksiewicz U. KRAB-ZFP Transcriptional Regulators Acting as Oncogenes and Tumor Suppressors: An Overview. *IJMS*. 2021;22(4):2212. doi:10.3390/ijms22042212
167. Busiello T, Ciano M, Romano S, et al. Role of ZNF224 in cell growth and chemoresistance of chronic lymphocytic leukemia. *Hum Mol Genet*. 2017;26(2):344-353. doi:10.1093/hmg/ddw427
168. Yu S, Ao Z, Wu Y, et al. ZNF300 promotes chemoresistance and aggressive behaviour in non-small-cell lung cancer. *Cell Prolif*. 2020;53(11). doi:10.1111/cpr.12924
169. Oguri T, Katoh O, Takahashi T, et al. The Krüppel-type zinc finger family gene, HKR1, is induced in lung cancer by exposure to platinum drugs. *Gene*. 1998;222(1):61-67. doi:10.1016/S0378-1119(98)00464-8

170. Yang L, Hamilton SR, Sood A, et al. The Previously Undescribed ZKSCAN3 (ZNF306) Is a Novel “Driver” of Colorectal Cancer Progression. *Cancer Res.* 2008;68(11):4321-4330. doi:10.1158/0008-5472.CAN-08-0407
171. Liu Y, Ouyang Q, Sun Z, et al. The Novel Zinc Finger Protein 587B Gene, ZNF587B, Regulates Cell Proliferation and Metastasis in Ovarian Cancer Cells in vivo and in vitro. *CMAR.* 2020;Volume 12:5119-5130. doi:10.2147/CMAR.S252347
172. Tian C, Xing G, Xie P, et al. KRAB-type zinc-finger protein Apak specifically regulates p53-dependent apoptosis. *Nat Cell Biol.* 2009;11(5):580-591. doi:10.1038/ncb1864
173. Huang C, Wu S, Li W, et al. Zinc-finger protein p52-ZER6 accelerates colorectal cancer cell proliferation and tumour progression through promoting p53 ubiquitination. *EBioMedicine.* 2019;48:248-263. doi:10.1016/j.ebiom.2019.08.070
174. Xing J, Zhang Y, Han K, et al. Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 2009;19(9):1516-1526. doi:10.1101/gr.091827.109
175. Sundaram V, Cheng Y, Ma Z, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* 2014;24(12):1963-1976. doi:10.1101/gr.168872.113
176. Lee A, Nolan A, Watson J, Tristem M. Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1626):20120503. doi:10.1098/rstb.2012.0503
177. Martin SL. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J Biomed Biotechnol.* 2006;2006(1):45621. doi:10.1155/JBB/2006/45621
178. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 2002;21(21):5899-5910. doi:10.1093/emboj/cdf592
179. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009;10(10):691-703. doi:10.1038/nrg2640
180. Cui F, Sirotin MV, Zhurkin VB. Impact of Alu repeats on the evolution of human p53 binding sites. *Biol Direct.* 2011;6:2. doi:10.1186/1745-6150-6-2
181. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 2008;9(5):397-405. doi:10.1038/nrg2337
182. Friedli M, Trono D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annu Rev Cell Dev Biol.* 2015;31(1):429-451. doi:10.1146/annurev-cellbio-100814-125514
183. White TB, Morales ME, Deininger PL. Alu elements and DNA double-strand break repair. *Mobile Genetic Elements.* 2015;5(6):81-85. doi:10.1080/2159256X.2015.1093067

184. Morales ME, White TB, Streva VA, DeFreece CB, Hedges DJ, Deininger PL. The Contribution of Alu Elements to Mutagenic DNA Double-Strand Break Repair. Jinks-Robertson S, ed. *PLoS Genet*. 2015;11(3):e1005016. doi:10.1371/journal.pgen.1005016
185. Doucet-O'Hare TT, Rodić N, Sharma R, et al. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci USA*. 2015;112(35). doi:10.1073/pnas.1502474112
186. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet*. 2020;52(3):306-319. doi:10.1038/s41588-019-0562-0
187. Kines KJ, Sokolowski M, deHaro DL, Christian CM, Belancio VP. Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Research*. 2014;42(16):10488-10502. doi:10.1093/nar/gku687
188. Griffin GK, Wu J, Iracheta-Vellve A, et al. Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. *Nature*. 2021;595(7866):309-314. doi:10.1038/s41586-021-03520-4
189. Deniz Ö, Ahmed M, Todd CD, Rio-Machin A, Dawson MA, Branco MR. Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat Commun*. 2020;11(1):3506. doi:10.1038/s41467-020-17206-4
190. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30. doi:10.3322/caac.21590
191. Küppers R. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer*. 2005;5(4):251-262. doi:10.1038/nrc1589
192. Zan H, Casali P. Epigenetics of Peripheral B-Cell Differentiation and the Antibody Response. *Front Immunol*. 2015;6. doi:10.3389/fimmu.2015.00631
193. Thieblemont C, Coiffier B. Lymphoma in older patients. *J Clin Oncol*. 2007;25(14):1916-1923. doi:10.1200/JCO.2006.10.5957
194. Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. *Blood*. 2018;131(21):2307-2319. doi:10.1182/blood-2017-11-764332
195. Moubadder L, McCullough LE, Flowers CR, Koff JL. Linking Environmental Exposures to Molecular Pathogenesis in Non-Hodgkin Lymphoma Subtypes. *Cancer Epidemiol Biomarkers Prev*. 2020;29(10):1844-1855. doi:10.1158/1055-9965.EPI-20-0228
196. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Sciences*. 2003;100(17):9991-9996. doi:10.1073/pnas.1732008100
197. Andor N, Maley CC, Ji HP. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Research*. 2017;77(9):2179-2185. doi:10.1158/0008-5472.CAN-16-1553

198. Jackson DA, Pombo A. Replicon Clusters Are Stable Units of Chromosome Structure: Evidence That Nuclear Organization Contributes to the Efficient Activation and Propagation of S Phase in Human Cells. *Journal of Cell Biology*. 1998;140(6):1285-1295. doi:10.1083/jcb.140.6.1285
199. Wang Y, Yang J, Wild AT, et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat Commun*. 2019;10(1):943. doi:10.1038/s41467-019-08905-8
200. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*. 2017;18(2):71-86. doi:10.1038/nrg.2016.139
201. Abrink M, Ortiz JA, Mark C, et al. Conserved interaction between distinct Krüppel-associated box domains and the transcriptional intermediary factor 1 beta. *Proc Natl Acad Sci U S A*. 2001;98(4):1422-1426. doi:10.1073/pnas.98.4.1422
202. Trono D. Transposable Elements, Polydactyl Proteins, and the Genesis of Human-Specific Transcription Networks. *Cold Spring Harb Symp Quant Biol*. 2015;80:281-288. doi:10.1101/sqb.2015.80.027573
203. Playfoot CJ, Duc J, Sheppard S, et al. Transposable elements and their KZFP controllers are drivers of transcriptional innovation in the developing human brain. *Genome Res*. 2021;31(9):1531-1545. doi:10.1101/gr.275133.120
204. Wagner S, Hess MA, Ormonde-Hanson P, et al. A Broad Role for the Zinc Finger Protein ZNF202 in Human Lipid Metabolism. *Journal of Biological Chemistry*. 2000;275(21):15685-15690. doi:10.1074/jbc.M910152199
205. Wang W, Shang W, Zou J, et al. ZNF667 facilitates angiogenesis after myocardial ischemia through transcriptional regulation of VASH1 and Wnt signaling pathway. *Int J Mol Med*. 2022;50(4):129. doi:10.3892/ijmm.2022.5185
206. Wang C, Ivanov A, Chen L, et al. MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation. *EMBO J*. 2005;24(18):3279-3290. doi:10.1038/sj.emboj.7600791
207. Tunbak H, Enriquez-Gasca R, Tie CHC, et al. The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. *Nat Commun*. 2020;11(1):5387. doi:10.1038/s41467-020-19170-5
208. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017;171(2):481-494.e15. doi:10.1016/j.cell.2017.09.027
209. Merdan S, Subramanian K, Ayer T, et al. Gene expression profiling-based risk prediction and profiles of immune infiltration in diffuse large B-cell lymphoma. *Blood Cancer J*. 2021;11(1):2. doi:10.1038/s41408-020-00404-0
210. Shannon M, Hamilton AT, Gordon L, Branscomb E, Stubbs L. Differential Expansion of Zinc-Finger Transcription Factor Loci in Homologous Human and Mouse Gene Clusters. *Genome Res*. 2003;13(6a):1097-1110. doi:10.1101/gr.963903

211. Lupo A, Cesaro E, Montano G, Zurlo D, Izzo P, Costanzo P. KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *CG*. 2013;14(4):268-278. doi:10.2174/13892029113149990002
212. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
213. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-273. doi:10.1038/ng1180
214. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
215. Larouche JF, Berger F, Chassagne-Clément C, et al. Lymphoma recurrence 5 years or later following diffuse large B-cell lymphoma: clinical characteristics and outcome. *J Clin Oncol*. 2010;28(12):2094-2100. doi:10.1200/JCO.2009.24.5860
216. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. In: von Stechow L, ed. *Cancer Systems Biology*. Vol 1711. Methods in Molecular Biology. Springer New York; 2018:243-259. doi:10.1007/978-1-4939-7493-1\_12
217. Olszewski AJ, Sharaf R, Marcus C, et al. *CDKN2A* Deletions Define an Unfavorable Subgroup within the *MYD88/CD79B* (MCD) Subtype of Diffuse Large B-Cell Lymphoma (DLBCL) and Are Mutually Exclusive with *TP53* mutations. *Blood*. 2021;138(Supplement 1):2392-2392. doi:10.1182/blood-2021-146335
218. Karube K, Enjuanes A, Dlouhy I, et al. Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets. *Leukemia*. 2018;32(3):675-684. doi:10.1038/leu.2017.251
219. Wright GW, Huang DW, Phelan JD, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell*. 2020;37(4):551-568.e14. doi:10.1016/j.ccell.2020.03.015
220. Vanlangenakker N, Vanden Berghe T, Vandenabeele P. Many stimuli pull the necrotic trigger, an overview. *Cell Death Differ*. 2012;19(1):75-86. doi:10.1038/cdd.2011.164
221. Kaya-Okur HS, Wu SJ, Codomo CA, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*. 2019;10(1):1930. doi:10.1038/s41467-019-09982-5
222. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science*. 2022;376(6588):44-53. doi:10.1126/science.abj6987
223. Lyu X, Chastain M, Chai W. Genome-wide mapping and profiling of  $\gamma$ H2AX binding hotspots in response to different replication stress inducers. *BMC Genomics*. 2019;20(1):579. doi:10.1186/s12864-019-5934-4

224. Sørensen CS, Syljuåsen RG. Safeguarding genome integrity: the checkpoint kinases ATR, CHK1 and WEE1 restrain CDK activity during normal DNA replication. *Nucleic Acids Res.* 2012;40(2):477-486. doi:10.1093/nar/gkr697
225. Macheret M, Halazonetis TD. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature.* 2018;555(7694):112-116. doi:10.1038/nature25507
226. Courtot L, Hoffmann JS, Bergoglio V. The Protective Role of Dormant Origins in Response to Replicative Stress. *Int J Mol Sci.* 2018;19(11):3569. doi:10.3390/ijms19113569
227. Hagan CR, Rudin CM. DNA cleavage and Trp53 differentially affect SINE transcription. *Genes Chromosom Cancer.* 2007;46(3):248-260. doi:10.1002/gcc.20406
228. Reisländer T, Lombardi EP, Groelly FJ, et al. BRCA2 abrogation triggers innate immune responses potentiated by treatment with PARP inhibitors. *Nat Commun.* 2019;10(1):3143. doi:10.1038/s41467-019-11048-5
229. Shen YJ, Le Bert N, Chitre AA, et al. Genome-Derived Cytosolic DNA Mediates Type I Interferon-Dependent Rejection of B Cell Lymphoma Cells. *Cell Reports.* 2015;11(3):460-473. doi:10.1016/j.celrep.2015.03.041
230. Naranbhai V, Viard M, Dean M, et al. HLA-A\*03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. *Lancet Oncol.* 2022;23(1):172-184. doi:10.1016/S1470-2045(21)00582-9
231. Calis JJA, Maybeno M, Greenbaum JA, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol.* 2013;9(10):e1003266. doi:10.1371/journal.pcbi.1003266
232. Larson AG, Elnatan D, Keenen MM, et al. Liquid droplet formation by HP1 $\alpha$  suggests a role for phase separation in heterochromatin. *Nature.* 2017;547(7662):236-240. doi:10.1038/nature22822
233. Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. *Nature.* 2017;547(7662):241-245. doi:10.1038/nature22989
234. Stewart-Morgan KR, Petryk N, Groth A. Chromatin replication and epigenetic cell memory. *Nat Cell Biol.* 2020;22(4):361-371. doi:10.1038/s41556-020-0487-y
235. Dubois S, Tesson B, Mareschal S, et al. Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles. *EBioMedicine.* 2019;48:58-69. doi:10.1016/j.ebiom.2019.09.034
236. Yang Y, Shaffer AL, Emre NCT, et al. Exploiting Synthetic Lethality for the Therapy of ABC Diffuse Large B Cell Lymphoma. *Cancer Cell.* 2012;21(6):723-737. doi:10.1016/j.ccr.2012.05.024

237. Zhang PP, Ding DZ, Shi B, et al. Expression of TRIM28 correlates with proliferation and Bortezomib-induced apoptosis in B-cell non-Hodgkin lymphoma. *Leukemia & Lymphoma*. 2018;59(11):2639-2649. doi:10.1080/10428194.2018.1452207
238. Gravett AM, N. Trautwein, Stevanović S, Dagleish AG, Copier J. Gemcitabine alters the proteasome composition and immunopeptidome of tumour cells. *OncImmunology*. 2018;7(6):e1438107. doi:10.1080/2162402X.2018.1438107
239. Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279-285. doi:10.1093/nar/gkv1344
240. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):D988-D995. doi:10.1093/nar/gkab1049
241. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360. doi:10.1038/nmeth.3317
242. Gentleman RC, Carey VJ, Bates DM, et al. [No title found]. *Genome Biol*. 2004;5(10):R80. doi:10.1186/gb-2004-5-10-r80
243. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29. doi:10.1186/gb-2014-15-2-r29
244. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015;43(7):e47-e47. doi:10.1093/nar/gkv007
245. Bryson T, Henikoff S. *3XFlag-PATn5 Protein Purification and MEDS-Loading (5x Scale, 2L Volume) V1.*; 2019. doi:10.17504/protocols.io.8yrhvxv6
246. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
247. Meers MP, Tenenbaum D, Henikoff S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & Chromatin*. 2019;12(1):42. doi:10.1186/s13072-019-0287-4
248. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*. 2019;9(1):9354. doi:10.1038/s41598-019-45839-z
249. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247
250. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
251. Ramírez F, Ryan DP, Grüning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-W165. doi:10.1093/nar/gkw257



252. Kara N, Krueger F, Rugg-Gunn P, Houseley J. Genome-wide analysis of DNA replication and DNA double-strand breaks using TrAEL-seq. Paull T, ed. *PLoS Biol.* 2021;19(3):e3000886. doi:10.1371/journal.pbio.3000886
253. Dodt M, Roehr J, Ahmed R, Dieterich C. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology.* 2012;1(3):895-905. doi:10.3390/biology1030895
254. Fu Y, Wu PH, Beane T, Zamore PD, Weng Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics.* 2018;19(1):531. doi:10.1186/s12864-018-4933-1
255. R: A language and environment for statistical computing. Published online 2013.
256. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics.* 2010;26(17):2204-2207. doi:10.1093/bioinformatics/btq351
257. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676-682. doi:10.1038/nmeth.2019
258. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat.* 2017;38(7):788-797. doi:10.1002/humu.23230
259. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 2020;48(W1):W449-W454. doi:10.1093/nar/gkaa379
260. Chong C, Marino F, Pak H, et al. High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferon-γ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Mol Cell Proteomics.* 2018;17(3):533-548. doi:10.1074/mcp.TIR117.000383
261. Chong C, Müller M, Pak H, et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun.* 2020;11(1):1293. doi:10.1038/s41467-020-14968-9
262. Pak H, Michaux J, Huber F, et al. Sensitive Immunopeptidomics by Leveraging Available Large-Scale Multi-HLA Spectral Libraries, Data-Independent Acquisition, and MS/MS Prediction. *Mol Cell Proteomics.* 2021;20:100080. doi:10.1016/j.mcpro.2021.100080
263. Bruderer R, Bernhardt OM, Gandhi T, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics.* 2015;14(5):1400-1410. doi:10.1074/mcp.M114.044305
264. Grandvaux N, Servant MJ, tenOever B, et al. Transcriptional Profiling of Interferon Regulatory Factor 3 Target Genes: Direct Involvement in the Regulation of Interferon-

Stimulated Genes. *J Virol.* 2002;76(11):5532-5539. doi:10.1128/JVI.76.11.5532-5539.2002

## **Filipe Martins**

Born on the 20<sup>th</sup> of July 1987 in Geneva

Address : 1, Avenue des Cavaliers / 1224 Chêne-Bougeries

# **CURRICULUM VITAE**

## **Academic Cursus :**

- Passed bilingual (German/French) « Maturité » diploma (scientific section) in 2006.
- **Medical doctor degree obtained in 2012 (Geneva, Switzerland-CH).**
- Passed the **ACLS exam (Advanced Cardiovascular Life Support)** in September 2014, HFR (Hôpital Fribourgeois, CH).
- **Post-graduate training in clinical research (co-investigator) at CHUV (Centre Hospitalier Universitaire Vaudois) in December 2014, Lausanne (CH).**
- Passed the modules : « *Techniques de communication centrées sur le patient atteint d'un cancer – Communication skills in cancer patient care* » in September 2016, Bern (CH).
- **Passed the specialty written exam of the European Society of Medical Oncology (ESMO) in October 2016, Olten (CH).**
- **Passed the specialty oral exam (FMH) of the Swiss Society of Oncology (SSOM) in October 2016, Olten (CH).**
- **Passed the Swiss Board Examination in General Internal Medicine in December 2017, Bern**
- **Passed the European Hematology Association (EHA) Examination in June 2018, Bern**
- **Passed the specialty written exam (FMH) of the Swiss Society of Hematology (SSH) in October 2018, Geneva (CH).**
- **Passed the specialty oral exam (FMH) of the Swiss Society of Hematology (SSH) in October 2018, Geneva (CH).**
- **Certificate of Advanced Studies in Personalized Medical Oncology (CAS-PMO) from the University of Basel and Lausanne in 2019.**
- **Doctorate of Medicine (MD) from the University of Lausanne (Unil, CH) in**

**February 2020.**

**Employment History :**

As of Apr2018: scientist in the **laboratory of virology and genetics of Professor Didier Trono at the School of Life Sciences, EPFL, Lausanne**, Switzerland. *Research field:* KRAB-ZNF proteins and Transposable elements (TEs) as regulators of human mature B-cell development; Research on their pleiotropic implications in solid tumors and hematological malignancies.

Nov2016-Mar2018: medical intern in the **oncology department of Professor George Coukos** and service of hematology of **Professor Michel Duchosal** at the University Hospital Centre of Vaud (CHUV) in Lausanne, Switzerland.

Nov2014-Oct2016: medical intern in the **oncology department of Professor George Coukos** and service of **Professor Solange Peters** at the University Hospital Centre of Vaud (CHUV) in Lausanne, Switzerland.

Dec2012-Oct2014: medical intern in the **internal medicine department of Professor Daniel Hayoz** at the Fribourg Cantonal Hospital. Performed several rotations during this period as follows:

- 6 months rotation in the **oncology department** of Professor Daniel Betticher from April to October 2014.
- 4 months rotation in the **emergency department** of Professor Vincent Ribordy from November 2013 to April 2014.
- 7 months rotation in the **service of nephrology/hemodialysis** under Dr. E. Descombes supervision from May 2013 to November 2013.

May2012–June2012: medical intern in the **internal medicine department of Professor Daniel Genné** at the Hospital of La Chaux-de-Fonds in Neuchâtel.

Prior to May 2012:

- Trainee in the **internal medicine department** of **Professor Arnaud Perrier** at the Geneva University Hospital (HUG), Switzerland, for two months in 2012.
- In early 2012, I was a trainee for one month in the **pediatric anesthesiology department** of **Professor Walid Habre** at the Geneva University Hospital (HUG), Switzerland.
- At two months trainee tour performed in December/January 2011-2012 in the **clinical pathology department** of **Professor Laura Rubbia-Brandt** at the University Medical Center (CMU) in Geneva, Switzerland.
- Two months as a trainee in the **department of infectious diseases** of **Professor Daniel Lew** at the Geneva University Hospital, Switzerland.
- One month as a trainee in the **intensive care unit** under **Professor Laurent Brochard** at the Geneva University Hospital, Switzerland.

**Parallel activity:**

- During 2010 and 2011, member of the Geneva section for the detection of melanoma skin cancer campaign called “Get to know your skin”.

**Thesis:**

Master thesis: “*The role of intensity modulation radiation therapy (IMRT) in the treatment of glioblastoma*” ; Director of thesis: **Professor Damien Weber** (Radiation Oncology department, Geneva University Hospital – HUG).

MD thesis: “*The combination of stereotactic radiosurgery with immune checkpoint inhibitors or targeted therapy in melanoma patients with brain metastases: a retrospective study*”, Director of thesis: **Professor Olivier Michielin** (Oncology department, University Hospital of Lausanne – CHUV).

## **Publications list:**

Hemett OM\*, Martins F\*, Descombes E\*, Betticher D, and Hayoz D. (2014). **Thrombotic microangiopathy: when time is the key factor.** *Rev Med Suisse*, 10, 794-803. \* Co-first authors.

Descombes E., Martins F.\*, Hemett O. M., Erard, V., & Chuard C. **Three-times-weekly, post-dialysis cefepime therapy in patients on maintenance hemodialysis: a retrospective study.** *BMC Pharmacol Toxicol.* 2016 Feb 4;17(1):4. doi:10.1186/s40360-016-0048-y.

Blum S., Martins F.\*, Alberio L. **Ruxolitinib in the treatment of polycythemia vera: patient selection and considerations.** *J Blood Med.* 2016 Sep 23 ;7 :205-215. eCollection 2016.

Martins F.\*, Berthold D. **Metastatic Renal Carcinoma, Back to the future.** *Rev Med Suisse.* 2016 Mai 18 ;12(519) :994-8.

Martins F.\*, Orcurto A., Michielin O., Coukos G. **Principles of adoptive cell therapy based on « Tumor Infiltrating Lymphocytes ».** *Rev Med Suisse.* 2016 Mai 18 ;12(519) :989-93.

Martins F., Pouw-Schoumns J., Georg Racila R., Spertini O., Blum S. **Characterization of myelodysplastic syndromes with transformation to acute lymphoblastic leukaemia.** Abstract EHA Mai 18, 2017; 182638 // **Abstract** : PB1924.

Blum S., Martins F.\*, Lübbert M. **Immunotherapy in adult acute leukemia.** *Leukemia Research.* June 2017.

Stadelmann R., Martins F., Blum S. **Immunotherapie bei acuter lymphatischer Leukämie.** *Schweizer Zeitschrift für Onkologie.* 2/2018.

Martins F.\*, Stalder G. **Multiple myeloma revealed by a solitary skin plasmacytoma.** *Clinics in Oncology.* 2018 Vol. 3 Article 1462

Voruz S., Martins F., Cairoli A., Naveiras O., Homicsko K., Missiaglia E., de Leval L., Bisig B., Michielin O., Blum S. **Comment On “MEK Inhibition With Trametinib And Tyrosine Kinase Inhibition With Imatinib In Multifocal Histiocytic Sarcoma”.** *Haematologica*. March 2018 103: e130

Martins F.\*, Stalder G., Obeid M. **Generating the Abscopal Effect by Combining Proapoptotic Peptides With IL-12-Based Immunotherapy.** *Neoplasia*. 2018 Feb;20(2):193-196.

Martins F.\*, Stalder G., Van Der Gucht A., Grandoni F., Cairoli A. **Intramuscular Follicular Lymphoma.** *Clin Nucl Med*. 2018 Sep;43(9):682-684.

Martins F.\*, Sykietis G.P., Maillard M., Montserrat F., Ribi C., Kuntzer T., Michielin O., Peters S., Coukos G., Spertini F., Thompson J. A., Obeid M. **New therapeutic perspectives to manage refractory immune checkpoint-related toxicities: personalized treatment algorithms beyond corticosteroids.** *Lancet Oncology* 2019; 20: e54–64.

Martins F.\*, Latifyan S., Sykietis G. P., Lamine F., Maillard M., Montserrat F., Shabafrouz K., Ribi C., Cairoli A., Guex-Crosier Y., Kuntzer T., Michielin O., Peters S., Coukos G., Spertini F., Thompson J. A., Obeid M. **Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance.** *Nature Reviews Clinical Oncology* (2019)

De Goycoechea D., Stalder G., Martins F.\* Duchosal MA. **Immune Checkpoint Inhibition in Classical Hodgkin Lymphoma: From Early Achievements towards New Perspectives.** *J Oncol*. 2019 May 7;2019:9513701

Martins F.\*, Schiappacasse L., Levivier M., Tuleasca C., Cuendet MA., Aedo-Lopez V., Gautron Moura B., Homicsko K., Bettini A., Berthod G., Gérard CL., Wicky A., Bourhis J., Michielin O. **The combination of stereotactic radiosurgery with immune checkpoint inhibitors or targeted therapy in melanoma patients with brain metastases: a retrospective study.** *J Neurooncol*. 2020 Jan; 146(1):181-193.

Martins F.\*, Kruszewski M, Scarpelli I, et al. **Characterization of myelodysplastic syndromes progressing to acute lymphoblastic leukemia.** *Annals of Hematology*. 2020 Jun. DOI: 10.1007/s00277-020-04114-2.

Martins F.\*, Obeid M. **Personalized treatment of immune-related adverse events — balance is required.** *Nature Reviews Clinical Oncology* (2020)

\*First or co-first authorships

### **Posters:**

Poster for the SSMI (Swiss Society of Internal medicine practitioners) in May 2014 and at the SGN-SSN society (Swiss Society of Nephrology) in December 2014 called “**Twice weekly post-dialysis Cefepime prescription in patients on maintenance hemodialysis**”. Martins F\*, Hemett OM, Erard V, CHuard C, Descombes E. Nephrology and infectious disease units. Department of internal medicine, HFR Fribourg Cantonal Hospital.

- Oral presentation at the 46<sup>th</sup> Annual meeting of the Swiss Society of Nephrology (December 2014).

Posters for the SGAIM/SSMIG in may 2016, Basel:

- **Sequential treatment with all five available tyrosine kinase inhibitors and successful re-exposure after recurrent pleural effusion to Dasatinib in an elderly patient with CML.** Filipe Martins\*, Olivier Spertini, Francesco Grandoni, Sabine Blum.
- **A rare presentation of a frequent disease: chronic lymphocytic leukaemia with prominent involvement of the central nervous system.** Francesco Grandoni, Mathilde Gavillet, Filipe Martins, Sabine Blum, Anne Cairolì

\*First or co-first authorship



**Spoken languages:**

English –Fluent

German – Fluent

Portuguese – Fluent

Spanish – Basic

**Hobbies :** Drawing, Art, Cinema