Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Armin Hadzic, MMedSc.

born in Rijeka, Croatia

Oral examination: 04.02.2021

Heterogeneity and multi-omics features of alternative lengthening of telomeres

Referees: Prof. Dr. Karsten Rippe

Dr. Sevin Turcan

Declaration

I hereby declare that I have written the submitted dissertation ”**Heterogeneity and multi-omics features of alternative lengthening of telomeres**” myself and in this process have used no other sources or materials than those explicitly indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

(Place, Date)

Armin Hadzic

Table of Contents

Summary

Telomere maintenance mechanisms are crucial for cancer cells as they are required for their unlimited proliferation capacity. While most cancers reactivate the reverse transcriptase telomerase, a significant fraction of tumors maintains telomere length without it. These cancers employ the alternative lengthening of telomeres (ALT) pathway, which relies on DNA repair and recombination to extend telomere repeats. ALT presence is primarily confirmed with the C-circle assay as gold-standard, which detects extrachromosomal telomere repeats that are only found in ALT-positive cells. Mutations within the repeat repressor ATRX/DAXX/H3.3 are overrepresented in ALT cancers. ALT presence is crucial for long-term survival in tumor entities where it is active. However, beyond the above-mentioned mutations, it is unclear which molecular features are key to understanding the mechanism by which ALT operates. This thesis addresses this issue with three aims: (i) Describing ALT activity heterogeneity in primary tumor samples; (ii) Using sequencing readouts to define molecular ALT features and extract a characteristic signature; (iii) Inhibiting epigenetic modifiers with drugs and observing their effect on viability in relation to ALT activity.

Firstly, I quantified C-circle levels in 687 primary tumor biopsies from sarcomas. The heterogeneous distribution indicates that ALT-activity can vary about tenfold within the same tumor entity. Next, I conducted ATAC-seq and RNA-seq of long and short RNAs in ALT positive and negative cell lines from pediatric glioblastoma and osteosarcoma to find shared ALT features. Information on open chromatin regions, transcriptome, miRNA, transposable elements, and piRNA was extracted from these data. From the ATAC-seq data, it was found that ALT+ cell lines had predominantly increased chromatin accessibility in non-coding regions. Binding motifs for AP-1 and RUNX transcription factors (TF) were enriched, whilst downregulated accessible regions result from reduced SOX TFs. The differential gene expression analysis revealed that immune TFs were enriched in upregulated ALT genes. This led to the identification of *NFATC2* as a potential ALT biomarker, as it was found in promoter regions of upregulated genes and through ATAC-seq based TF motif analysis. The immune-related TFs may be affected by the genetic instability inherent to ALT. However, the multi-omics ALT signature indicated that the cell lines have a reduced response to oxidative stress. These factors may cooperate in inducing a heightened inflammatory state that drives chromatin accessibility and gene expression. Differential miRNAs were extracted and could explain both TERT and SOX downregulation and RUNX upregulation, indicating another gene regulatory mechanism employed by ALT cell lines. Furthermore, an integrative multi-omics analysis was performed to extract an ALT signature, which was governed by gene expression, miRNA, and chromatin accessibility. As more upregulated open regions in ATAC data were observed, inhibitors for EZH2 that sets the repressive histone H3 lysine 27 trimethylation (H3K27me3) and DNA methylases were correlated with an ALT-specific lethality and survival, respectively. Another aberrant epigenetic feature found in ALT, an H3.3S31p chromosome-wide signal during mitosis, was studied with different inhibitors. The kinase HASPIN was found to reduce aberrant H3.3S31 phosphorylation upon treatment with a corresponding inhibitor. This kinase is involved in chromosomal segregation and links ALT genetic instability to DNA damage signaling during mitosis.

In summary, the findings from this thesis reveal that ALT activity is heterogeneous in primary tumor cells. The analysis with different omics readouts points to significant differences in the transcriptome, chromatin accessibility, and miRNA expression and yields a multi-omics signature to identify ALT presence. Lastly, the correlation between cell viability and ALT-activity upon treatment with EZH2 and DNA methylase inhibitors and the HASPIN mediated aberrant H3.3S31p signal during mitosis point to ALT-specific epigenetic features. These could be further exploited in preclinical studies for patient stratification and identification of novel drug targets specific for ALT.

Zusammensfassung

Telomer-Instandhaltungs-Mechanismen sind kritisch für Krebszellen, da sie für deren unbegrenzte Proliferationskapazität nötig sind. Die meisten Krebszellen halten ihre Telomere durch Reaktivierung der reversen Transkriptase Telomerase instand. Ein signifikanter Teil der Tumoren ist jedoch nicht auf Telomerase angewiesen und nutzt stattdessen alternative, telomerverlängernde Mechanismen (ALT). Diese basieren auf DNA-Reparatur und Rekombinantionsprozessen. Die Quantifizierung von extrachromosomalen Telomersequenzen durch den „C-circle assay“ dient als Goldstandard-Nachweis für das Vorhandensein von ALT in Krebszellen. ALT-positive Tumore zeigen überproportional häufiges Auftreten von Mutationen in Proteinen des ATRX/DAXX/H3.3-Komplexes. Es wird vermutet, dass diese Mutationen mit Funktionsstörungen der Telomere assoziiert sind. Obwohl der ALT-Status eines Tumors die Langzeitüberlebensrate beeinflusst, ist es zum gegenwärtigen Zeitpunkt nicht möglich, ALT durch Analyse von „Omics“-Daten zu erkennen. Während in den letzten Jahrzehnten viele Fortschritte gemacht wurden, sind auch die molekularen Grundlagen von ALT noch immer nicht bekannt. Diese Dissertation beleuchtet diese offenen Fragen aus drei verschiedenen Richtungen: i) Analyse und Beschreibung der ALT-Heterogenität in primären Tumorproben; ii) Beschreibung von ALT auf Grundlage von Sequenzierungsdaten und Entwicklung einer ALT-Signatur; iii) Inhibition von epigenetischen Modifikationsmechanismen und Beschreibung des Effektes auf das Überleben von ALT-Zellen.

Ich habe zunächst die „C-Circles“ in 687 primären Sarkombiopsien quantifiziert. Ich konnte zeigen, dass die Menge der C-Circles in den Tumorproben einer heterogenen Verteilung folgt und sich um bis zu zehnfach innerhalb derselben Tumorart unterscheidet. Anschließend habe ich ALT-positive und ALT-negative Glioblastom- und Osteosarkom-Zelllinien über ATAC-seq und RNA-seq auf gemeinsame ALT-Charakteristika untersucht. Informationen zu offenen Chromatinregionen, Transkriptom, miRNAs, transposablen Elementen, und piRNAs wurden aus diesen Daten extrahiert. Die ATAC-seq Daten zeigten, dass ALT-positive Zelllinien generell erhöhte Chromatinzugänglichkeit in nicht-kodierenden Regionen des Genoms zeigen. Dabei waren Bindungsmotive der Transkriptionsfaktoren (TF) AP-1 und RUNX gehäuft, Motive der SOX TFs wurden dagegen weniger häufig als erwartet gefunden. Die Transkriptomananalyse zeigte, dass Immun-TFs eine wichtige Rolle in den beobachteten Genexpressionsveränderungen spielen. Daraus folgte die Identifikation des TF

NFATC2 als möglichem ALT-Biomarker. Bindungsmotive von NFATC2 wurden in Promoterregionen von hochregulierten Genen gefunden. Weiterhin wurde NFATC2 in der ATAC-seq Motivanalyse gefunden.

Genetische Instabilität, ein fundamentales ALT-Charakteristikum, könnte die bereits erwähnten Immun-TF beeinflussen. Die integrierte „Omics"-Analyse zeigte jedoch, dass ALT-positive Zelllinien eine verringerte Reaktion auf oxidativen Stress zeigen. Diese verschiedenen Faktoren könnten zusammen für den erhöhten Entzündungsstatus dieser Zellen verantwortlich sein und als Folge Chromatinzugänglichkeit und Genexpression beeinflussen.

Die Analyse von miRNA Expression konnte sowohl die TERT- und SOX-Herunterregulierung, als auch die RUNX-Hochregulierung bestätigen. Dies deutet auf einen weiteren genregulatorischen Mechanismus von ALT Zellen hin. Zusätzlich zu den beschriebenen Einzelanalysen wurde eine integrative „Multi-Omics" Analyse durchgeführt mit dem Ziel, eine ALT-Signatur zu extrahieren. Die gefundene Signatur beinhaltete Daten aus Genexpression, miRNAs, und Chromatinzugänglichkeit.

Da in der ATAC-seq Analyse mehr zugängliche Chromatinregionen in ALT gefunden wurden, habe ich Inhibitorwirkungen gegen EZH2, das für die repressive H3 Lysin 23 Dreifachmethylierung verantwortlich ist, und Methylasen mit ALT-spezifischen Effekten auf Zellviabilität korreliert. Weiterhin habe ich ein anderes epigenetisches Charakteristikum von ALT, das chromosomenweite Auftreten von H3.3S31p während der Mitose, mit verschiedenen Inhibitoren untersucht. Meine Experimente konnten zeigen, dass die Kinase HASPIN für das erwähnte H3.3S31p Signal verantwortlich ist. Diese Kinase ist Teil des Chromosomentrennungsmechanismus und verknüpft daher genetische Instabilität in ALT mit Signalwegen, die während der Mitose durch DNA-Schäden aktiviert werden.

Die Erkenntnisse dieser Dissertation zeigen, das ALT-Aktivität in primären Tumorzellen heterogen ist. Die „Multi-Omics" Analyse zeigt signifikante Unterschiede bezüglich Transkriptom, Chromatinzugänglichkeit, und miRNA Expression auf. Weiterhin ergibt sich eine ALT-Signatur, mit der sich ALT aufgrund der vorhergenannten Daten identifizieren lässt. Die Korrelation zwischen Viabilität und ALT-Aktivität nach EZH2- und DNA-Methylase-Inhibition, sowie die HASPIN-abhängigen Änderungen im H3.3S31p Signal während der Mitose deuten auf ALT-spezifische, epigenetische Charakteristika hin. Diese könnten in zukünftigen präklinischen Studien zur Patientenstratifikation und Entdeckung neuer, ALT-spezifischer pharmazeutischer Wirkungsmechanismen genutzt werden.

List of Abbreviations

- Pol	Phi29 DNA polymerase negative
+ Pol	Phi29 DNA polymerase positive
a.u.	Arbitrary units
aa	Amino acids
ADD	ATRX-DNMT3-DNMT3L
ALT	Alternative lengthening of telomeres
APB	ALT-associated PML body
ATR-X	X-linked alpha-thalassemia/mental retardation
BIR	Break-induced replication
BrdU	Bromodeoxyuridine
C-circles	C-rich extrachromosomal repeats
DIPG	Diffuse intrinsic pontine glioma
D-loop	Displacement loop
DMEM	Dulbecco's modified eagle medium
EC_{50}	Effective concentration 50 %
IC_{50}	Inhibitory concentration 50 %
ECTR	Extrachromosomal telomere repeat
EST	Ever shortening telomere
FBS	Fetal bovine serum
FC	Fold change
FDR	False discovery rate
FISH	Fluorescence *in situ* hybridization
FRiP	Fractions of reads in peaks
GO	Gene ontology
HR	Homologous recombination
IF	Immunofluorescence
ITS	Internal telomere sequence
lncRNA	Long noncoding RNA
MiDAS	Mitotic DNA synthesis
MOFA	Multi-omics factor analysis
OS	Osteosarcoma
PADOG	Pathway analysis with down-weighting of overlapping genes
PCA	Principal component analysis
pGBM	Pediatric glioblastoma
qPCR	Quantitative polymerase chain reaction
RIN	RNA integrity number
R-loops	DNA:RNA hybrids
RPM	Revolutions per minute
RT	Room temperature
S.D.	Standard deviation
SEM	Standard error of the mean
snoRNA	Small nucleolar RNA
TE	Transposable elements

TF	Transcription factor
T-loop	Telomere loop
TMM	Telomere maintenance mechanism
TPE-old	Telomere position effect over long distance
UMAP	Uniform manifold approximation and projection
WT	Wild type

Introduction

1 Telomere structure and maintenance

Telomeres, the ends of the linear chromosomes, need to be protected from being recognized as a double-stranded break by the cell's DNA repair machinery. This is avoided by shelterin proteins, which hide the exposed 3' end of the telomere into an upstream telomeric repeat and forms the T-loop (1). Another impact of linear genomes is the end-replication problem (2). Due to the necessity of RNA primers for replicating the lagging strand, the most distal 3' end will not be replicated entirely. Dividing cells, therefore, experience 50-200 bp telomere loss upon each completed replication cycle (3). As a cell continues to divide, telomeres will reach a point where they cannot form T-loops, shelterin proteins lose their binding to telomeres, and DNA repair processes will recognize the exposed 3' end. This forces the cells into a replicative crisis, whereby senescence and cell death processes will keep this cell from proliferating (4). Hence, differentiated somatic cells experience progressive telomere erosion, forcing somatic cells to rely on telomere elongation during embryogenesis by the telomerase holoenzyme (5). Some cells can incur mutations to escape replicative crisis and may further progress into cancer. To this end, most cancers simply reactive telomerase and maintain replicative immortality (6). Still, it was noted early in the 1990s that some cancers had no detectable telomerase activity and employed a telomere maintenance mechanism (TMM) that was eventually termed the alternative lengthening of telomeres (ALT) (7). As our collective knowledge increased, it was established that ALT is a complex phenomenon through which telomere stress and DNA repair cooperate in telomere extension (8).

1.1 Telomeres in a physiological setting

Human telomere sequences consist of repeats of the TTAGGG hexamer, where telomere lengths in newborn skin and blood cells are around 10 kb (9). The telomere length decreases to around 7 kb in similar tissues in elderly people (10). Due to the end-replication problem, telomeres contain a single-stranded 3' overhang (2). The shelterin complex conceals this exposed end by first forming a partial strand separation in the so-called displacement-loop (D-loop) upstream of the 3' overhang (Figure 1)(1). As the telomere folds back into a T-loop, the exposed 3' invades the double-stranded telomere repeat regions and form a D-loop stabilized by the shelterin proteins (1). This complex is comprised of TRF1 and TRF2, which bind the double-stranded telomeric sequences. In addition, POT1 binds the single-stranded 3' telomeric overhang, followed by recruitment of TIN2 and TPP1, which binds TRF1 and TRF2. Lastly, the Rap1 protein stabilizes the complex by binding to TRF2, and the T-loop is formed by linking all these components (1). The shelterin complex is essential to cell viability, and without its function, each telomere end would activate the DNA damage kinases ATM/ATR (11). This is seen in cells undergoing replicative crisis, where the telomere reaches a critically short length, at which point T-loop formation is impossible.

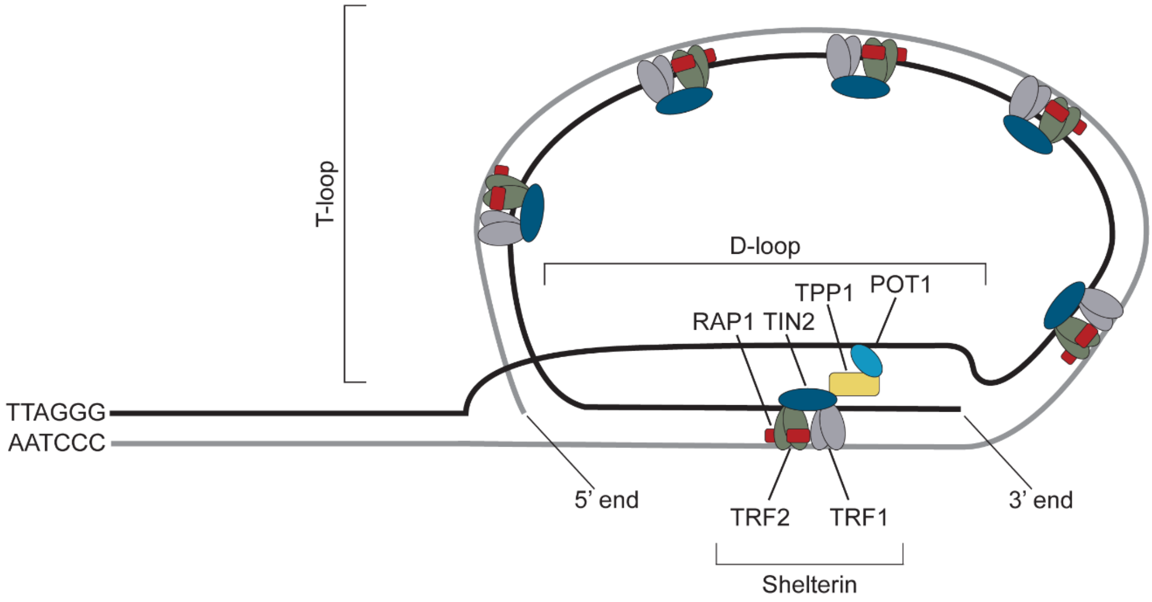


Figure 1. Shelterin proteins bridge telomere 3' overhangs to avoid activation of DNA damage sensing pathways. The shelterin proteins form the displacement loop (D-loop), which the exposed 3' telomere end is concealed into. The entire structure of the D-loop and subsequent telomere strand looping is called the T-loop.

The replicative crisis is believed to occur when telomeres shorten below 4-6 kb (12). When the exposed telomere is recognized by DNA damage signaling pathways, the cells undergo apoptosis or senesce via p53 and RB (13). Like all repetitive elements, telomeres are genetically fragile sites, meaning they are susceptible to DNA damage (14). Furthermore, non-B-DNA structures called G-quadruplexes can form from the G-rich telomere strand, which impedes the replisome and causes replication fork collapse. This is resolved by numerous specialized helicases, such as WRN, BLM*,* and ATRX (15-17), and in their absence, failure to solve replication fork collapse leads to DNA damage (18). Telomeres may also be more sensitive to DNA base damage, such as the formation of 8-oxoguanine from reactive oxygen species, although it is unknown why (19). The shelterin proteins mediate telomeric DNA repair, which strictly orchestrates specific DNA repair pathways depending on the DNA damage in question (20). The shelterin proteins are also believed to actively inhibit non-homologous end joining to avoid accidental telomere end-joining (21), and show a general preference towards homologous recombination (HR), and excision repair pathways (22).

1.2 Telomere extension in human cells

Telomere elongation mainly occurs during embryogenesis in embryonal stem cells, where the reverse transcriptase telomerase is activated and extends the telomere repeats (23). Telomerase consists of a protein component called TERT, an RNA-component called TERC, and the DKC1 protein together with three small nucleolar RNA (snoRNA) (NHP2, NOP10, GAR1) (24). TERT functions as the reverse transcriptase, TERC as an RNA component that serves as a template, and the DKC1 protein with snoRNAs stabilizes telomerase and enhances DNA synthesis (18). Telomerase can only add around 50 bp telomere repeats per strand and cell cycle (25, 26), and thus any telomere erosion occurring faster than this will result in a net loss. For example, adult stem cells are distinct from embryonal stem cells in that they show gradual aging and exhaustion despite active telomerase (27). Additionally, other cells show signs of telomere extension, such as keratinocytes, germ cells, and actively proliferating B and T cells (28)**.** Again, the active telomerase does not result in a net gain of telomere length (29). Therefore, it is likely that mesenchymal stem cells and cancer cells are the only known entities that show a net gain of telomeres to sustain their proliferation (28, 30).

1.3 Epigenetic state of the telomere

Telomeres were initially thought to be in a heterochromatic state, with an enrichment of the repressive histone marks H3K9me3 and H4K20me3 (31). Furthermore, other telomere features, such as high nucleosome density (32), compact chromatin (33), and that repetitive elements are generally repressed (34), pointed towards a heterochromatic state. Nevertheless, telomeric chromatin may be much more dynamic than previously thought. A recent study was able to differentiate between internal telomere sequences (ITS) from actual telomere repeats and compared ChIP-seq data from primary and embryonal cells. ITSs are telomeric repeats outside the telomere and are considerably enriched in the telomere adjacent regions called subtelomere. These telomeric sequences, however, may be differently regulated than actual telomeres (35). By distinguishing between ChIP signals from telomeres and ITS more carefully, the authors found a bivalent state of telomeric chromatin, comprising both euchromatic and heterochromatic marks (36). As such, it is worth noting that the epigenetic state of physiological telomeres is still debated. Since telomeres are more euchromatic during pluripotency (37), it is duly possible that the epigenetic state of telomeres depends on cell-specific circumstances. Future research built on telomeric capture strategies from the above-mentioned study will undoubtedly be informative in settling this question.

1.4 The relationship between subtelomere and telomere

The telomere sequences do not abruptly stop at a given locus, but rather the telomeric repeats are gradually more interspersed with other genetic elements (including coding genes such as *ISG15*) in a transition state called the subtelomere (35). The subtelomere is a genetically diverse region, both between different chromosomes and individuals (35). The subtelomere is also associated with innate immunity. For example, *ISG15* (an interferon-induced protein) is upregulated upon telomere shortening and causes inflammation (38). This indicates that short telomeres may induce pathways that are independent of T-loop loss. The process of telomere length affecting the subtelomeric region is called telomere position effect over long distance (TPE-OLD) (39), and consists of the telomere looping into the subtelomeric region. *TERT* expression is also connected to TPE-OLD, where again long telomeres repress *TERT,* and short telomeres allow its expression. This has been implicated explicitly in *TERT* re-expression in actively proliferating T-cells (40) and could prove an essential mechanism for *TERT* expression in other tissues and cancer. Another critical feature of subtelomeres is the high levels of CpG DNA methylation, the extent of which is ultimately associated with telomere length (41). Studies in patients with ICF syndrome, an immunodeficiency and developmental disease caused by mutations within *DNMT3B* and related DNA methylation proteins, have been enlightening in this case. These patients have hypomethylated subtelomeres and short telomeres (42). The telomere shortening mechanism has not been fully elucidated, but it is noted that high expression of the long non-coding RNA (lncRNA) TERRA may be involved (43). TERRA is transcribed from the subtelomeric region and may span into the telomeric sequences as well. Therefore, its length varies between 100 bp to 9 kb and includes telomeric and non-telomeric sequences (44). TERRA has been further implicated in having a diverse role in telomere maintenance (45), and can negatively regulate telomere length by forming telomeric R-loops, which cause replication stress (46, 47).

1.5 ATRX/DAXX/H3.3 in a physiological setting

The best-known mutations in ALT cancers are within the ATRX/DAXX/H3.3 deposition pathway (84). These proteins are instrumental in suppressing transcription from heterochromatic regions (85). While it is known that this complex is also associated with telomeres, its functions there are less clear. As described above, the heterochromatic state of the physiological telomere is debated, but it is clear that disrupting this pathway leads to telomere dysfunction (86-88). ATRX is a multifunctional chromatin remodeler of 260 kDa molecular weight (89). It was initially identified as commonly mutated within the rare X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome (90). This syndrome consists of a complex disease spectrum, including intellectual disabilities, alpha-thalassemia, and developmental defects. ATRX mutations in this disease were mainly found in the ATRX-DNMT3-DNMT3L (ADD) and helicase domains (Figure 3).

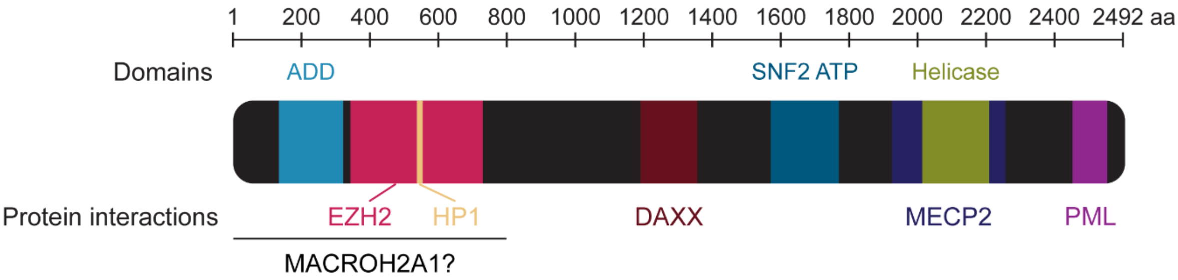


Figure 2. ATRX domains and protein interaction regions.

The ADD domain is believed to be responsible for reading H3K4me2 and H3K9me3, where the latter interaction is essential for targeting ATRX to heterochromatic compartments in the genome in the absence of H3K4 methylation (91). Mutations within the helicase domain have been linked to resolving G-quadruplexes, as evidenced by alpha-thalassemia and mutated ATRX. In this case, G-rich sequences above the α-globin gene can form G-quadruplexes that impede this gene's expression (92, 93). ATRX contains EZH2 and HP1 binding domains that are important for maintaining facultative and constitutive heterochromatin (84, 94). The EZH2 binding site has proven crucial for X-inactivation, where ATRX binds both the lncRNA XIST and the PRC2 complex to form facultative heterochromatin on the inactive X-chromosome (95, 96). The HP1 binding site is believed to target ATRX to heterochromatin (97). Within the N-terminal region of ATRX, there is an additional tentative binding site to a histone variant called macroH2A (98); however, in this instance, ATRX is believed to regulate macroH2A on G-rich regions negatively. Although the exact mechanism is unclear, replacing macroH2A from telomeres with H3.3 may resolve G-quadruplexes (99). The chromatin remodeling activities of ATRX lie in the SNF2 ATPase domain and helicase domain (84). These cooperate with the abovementioned protein interaction sites to properly integrate H3.3 into chromatin, and the latter domain is believed to confer identification and binding to G-quadruplexes (100). Lastly, towards the C-terminal end of the protein, two interaction sites for MECP2 and PML were identified, respectively (101, 102). MECP2 is crucial for neuronal development, and its general function concerns repressing genes with DNA methylation and is also involved in X-inactivation (103, 104). The MECP2 binding function of ATRX has been explicitly implicated in causing intellectual disabilities for patients with ATR-X syndrome (89). In this instance, lack of ATRX-MECP2 binding may lead to dysfunction in neural progenitors and thus affects neurodevelopment (105). The interaction site with PML connects ATRX to various processes, including virus repression (106, 107) and heterochromatin function (108). In the middle of the ATRX protein lies the DAXX interaction domain, which is essential in depositing H3.3 (84).

DAXX was initially identified as a FAS receptor ligand (109). Its function has since been implicated in many different pro-apoptotic processes, where one finding was that DAXX is recruited to the PML nuclear body to promote apoptosis (110). Still, its role as an H3.3 chaperone is the most important in this context (111). H3.3 is a histone variant that only differs from canonical H3 by 4/5 amino acids and is expressed from the *H3F3A* and *H3F3B* locus irrespective of replication (112). H3.3 is deposited onto euchromatic regions via HIRA and to heterochromatic and telomeric regions via ATRX/DAXX (112) (see Figure 4). The former process is critical for actively transcribed genes and regulatory elements and the latter for repressing heterochromatic regions and maintaining telomere stability (113).

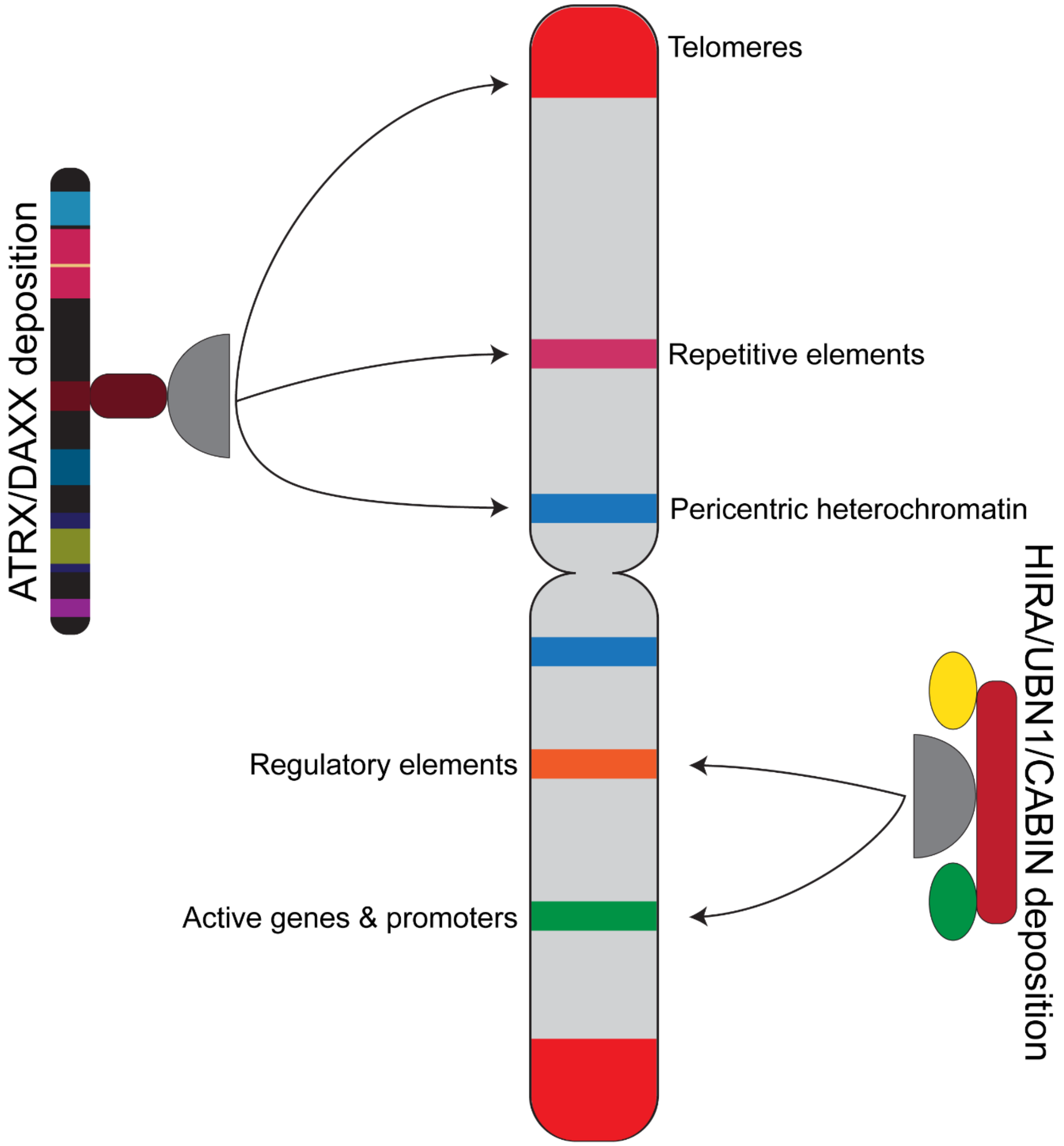


Figure 3. Deposition of the H3.3/H4 dimer by the ATRX/DAXX proteins onto repetitive regions and by the HIRA/UBN1/CABIN proteins onto regulatory elements and active genes and promoters. Figure adapted from (112).

As previously mentioned, H3.3 shows a high degree of homology with canonical H3. The unique S31 residue can be phosphorylated, whilst the other amino acids are chaperone motifs significant for DAXX or HIRA binding (112). During mitosis, H3.3S31p is localized to the pericentromeric region (114). This phosphorylation may have various functions, such as inducing gene expression pathways in cell differentiation (115), resolving UV damage during replication (116), and inducing cell death when chromosomes are missegregated (117).

2 Telomere maintenance mechanisms in cancer

2.1 Cancer telomere maintenance mechanisms

Most cancer cells employ telomere maintenance mechanisms (TMM) to elude replicative crisis. Approximately 90 % of all cancer entities show signs of telomerase activity, and about 10 % use the ALT mechanism (48). An exception is the ever-shortening telomere (EST) phenotype found in some neuroblastomas (49), which can spontaneously regress due to lack of TMM (50, 51). ALT is overrepresented in cancers of mesenchymal origin, namely soft tissue, nervous system, and bone cancers (48). It is still unclear why, as previously stated, mesenchymal stem cells are the only somatic cells that have a net gain TMM *via* telomerase (28, 30). However, this could indicate that these cells are not constrained by specific differentiation processes occurring in other cancers (48). ALT presence has a profound impact on long-term survival depending on the tumor entity. For example, ALT-positive neuroblastomas show low overall survival irrespective of initial tumor grading (52). On the other hand, ALT adult gliomas show a more favorable prognosis (53). As such, ALT presence can confer prognostic capabilities depending on the tumor entity. Intriguingly, immortalized cell lines show a high degree of ALT as well (54). There appears to be no common link between immortalization strategies, e.g., SV40 transduction or spontaneous transformation.

2.2 Telomere maintenance mechanism as a therapeutic target

The TMM has long been thought of as a golden-bullet strategy for specific cancer drugs. Telomerase represents a specific target that can be inhibited by various means, yet so far, no anti-telomerase drug has excited clinical trials successfully (55). There may be various reasons, such as the presence of telomerase in, e.g., stem cells and keratinocytes lowering therapeutic tolerance for patients and noncanonical functions of telomerase itself (56). There is also additional concern that targeting telomerase may instead lead to the emergence of ALT cancer populations. It has been shown that both TMMs can coexist within one given cancer population *in vitro* (57, 58) and *in vivo* (59). As such, any TMM drug would need to target both telomerase and ALT. However, druggable targets of ALT are currently lacking. Even if one could successfully interrupt the TMM, it is questionable if these cancers would reach replicative crisis before killing the patient. A back-of-the-envelope calculation of a 10 kb telomere shows that the cells can continue growing for about 60 population doublings. This assumes that crisis occurs at 4 kb telomere length, with telomere attrition at 100 bp per division. Assuming no cells die, $5.5 \cdot 10^{17}$ cells can be formed from one clone with no active TMM. Assuming cells weigh one nanogram (60), this would create a theoretical maximum of a 500,000-kilo tumor from one single cell. While this is obviously an oversimplification, it shows that even if the TMM is inhibited, the cancers may grow long before entering replicative crisis. In parallel, neuroblastomas that regress spontaneously are not wholly dependent on short telomeres but rather work in conjunction with immune clearance, growth factor deprivation, and DNA methylation (51). Taking inspiration from this, any TMM therapeutic would most likely be a combinatorial treatment. For example, by inhibiting both TMMs and using compounds that actively shorten telomeres in fast-dividing cells. The latter option could consist of G-quadruplex stabilizers that create DNA damage in both ALT and telomerase-positive cancers (61, 62). However, to get to this stage, we must further our knowledge in ALT to find druggable targets or susceptibilities inherent to ALT.

2.3 ALT mechanistic details

The ALT mechanism is dependent on replication stress which drives recombination (63). An overview of all of the ALT hallmarks and their biological background is seen in Table 1. Mutations within ATRX/DAXX, the introduction of telomere variants, and R-loops from TERRA exacerbate the replication stress already inherently present in telomeres (63). Mutations within the ATRX/DAXX/H3.3 deposition axis will be discussed further in section 2.4 (75, 76). Telomere variants, such as TCAGGG and TGAGGG, are introduced by mutagenesis and are a common feature in ALT (63, 77). They reduce shelterin binding and might elicit aberrant DNA repair processes that would typically be suppressed. The variants can also directly mediate binding by nuclear receptors, which aggravates replication stress (78). TERRA also contributes to replication stress, perhaps by forming DNA:RNA hybrids (R-loops) which hinder polymerases and DNA repair proteins (46).

Table 1. Hallmarks of ALT.

Hallmark	Readouts	Biological background
Aberrant H3.3S31p	H3.3S31p IF	Chromosome-wide staining of the usually centromeric mitotic H3.3S31p. Unclear background (64)
APB	IF of PML colocalization with telomere FISH signal	Proposed active sites of telomeric recombination (65)
C-circles	Amplification of C-rich extrachromosomal telomere repeats	Proposed byproduct and/or template for ALT telomere extension (48, 66)
G-quadruplexes	Fluorescent G-quadruplex ligands/BG4 antibody at telomeres	Induces replication stress, especially in the ALT context (17, 67)
High TERRA levels	TERRA FISH signal or qPCR	May be induced by short telomeres (45)
Long telomeres	Telomere signal from FISH, qPCR, or sequencing	General features of ALT recombination (68)
Non-S Phase DNA synthesis	IF of arrested cells with BrdU, amongst other specialized methods	BIR/MiDAS is active outside S-phase (69, 70)
Telomere dysfunction induced foci	IF of 53BP1 and γH2AX with telomere FISH	DNA damage at telomeric foci (71)
Telomere sister chromatid exchange	Chromosome oriented telomere FISH	A general feature of ALT recombination (72)
Telomere variants	Sequencing, variant FISH probes	Mutagenic byproduct possibly from BIR (73, 74)

It was recently shown that two distinct homologous recombination (HR) mechanisms are present in ALT cancer cells (Figure 2). These consist of break-induced replication (BIR) pathway, which is mainly active in G2/M (79), and the mitotic DNA synthesis (MiDAS) pathway during mitosis (80, 81). The BIR pathway has been known in yeast model systems for decades, though its role in telomeric human DNA repair was only recently elucidated (79). As the telomere experiences double-stranded breaks from replicative stress, it likely undergoes recombination with other telomeres in specialized nuclear compartments termed ALT-associated PML bodies (APBs). The APBs contain factors critical for HR and DNA damage response (e.g., RAD52) (69, 82).

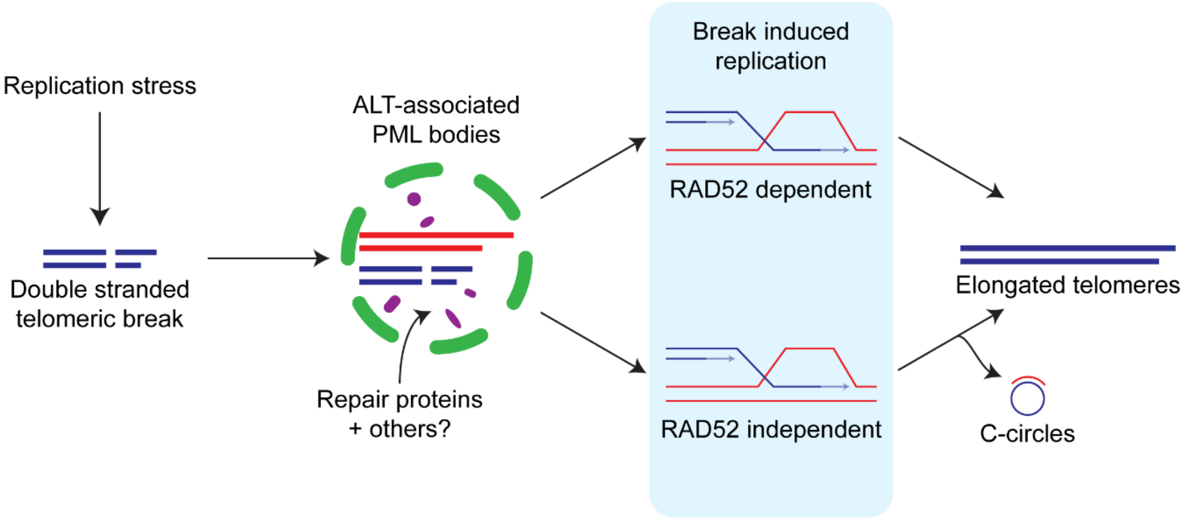


Figure 4. Possible mechanism of ALT telomere recombination. Adapted from (69).

The damaged telomere is elongated via BIR using the healthy telomere as a template. ALT BIR consists of two processes that are either dependent on RAD52 or independent, where the latter process is believed to generate circular C-rich extrachromosomal telomere repeats (C-circles). C-circles are currently the best quantifiable markers of ALT-activity (66), and as such, the RAD52 independent pathway may be more critical. An alternative model for the generation of C-circles is the sporadic release of telomere-internal single-stranded loops upon damage or telomere processing (83). It is not known whether both BIR and MiDAS can drive ALT recombination, if at all. One functional consequence of BIR, which is more mutagenic than other HR pathways (74), could be telomere variants. While many questions remain regarding the ALT recombination mechanism, it is evident that many different factors cause replication stress which then drives DNA damage at telomeres. The damage is resolved by different HR pathways, which in turn elongate the ALT telomere.

2.4 ATRX/DAXX/H3.3 mutations in ALT

ATRX and, to a lesser extent, DAXX are recurrently mutated in ALT cancers (84). However, their many functions make it difficult to answer why they act as ALT suppressors. ATRX/DAXX mutations generally tend to lead to loss of expression (84). Whilst DAXX mutations are only found in a specific tumor entity, ATRX mutations are found throughout ALT cancers (84, 109, 118). As such, research has been mainly focused on the connection between ALT and ATRX. Studies elucidating this link saw that loss of ATRX by itself causes extensive genome instability (119, 120). Reintroducing functional ATRX in ALT cell lines leads to suppression of ALT, perhaps by resolving G-quadruplexes and replication stress (93). Another study indicated a central role in replacing macroH2A with H3.3 at telomeres, whereby macroH2A accumulation causes replication stress and DNA damage (121). Therefore, it appears that ATRX disruption drives the replication stress that is fundamental to ALT extension.

H3.3 is a common mutation in pediatric brain cancers, and these mutations are almost entirely found in the same cancers (76). Therefore, its relevance to ALT is unclear. The two H3.3 mutations concern amino acid substitutions at K27 and G34, located in the histone tail (122). The first mutation leads to a methionine (M) substitution in K27 and profoundly alters the global H3K27me3 distribution (123). While the exact mechanism is still debated, especially concerning ALT, the K27M mutant probably sequesters the PRC2 complex that forms the repressive H3K27me3 (123, 124). Thereby, introducing the K27M mutant changes gene expression associated with many different cancer pathways (123, 125). The second mutation encodes an amino acid switch from G34 to lysine (K) or arginine (R). The G34V/R substitution is less characterized, but it is probable that it restricts neighboring H3K36me3 by inhibiting SETD2 binding (126). Some evidence also suggests that G34V/R inhibits PRC2 binding to H3K27 in a manner distinct from K27M mutants (127). Altogether, both H3.3 mutants change the epigenetic landscape that appears to be integral for driving pediatric brain cancers.

3 Scope of the thesis

Telomere maintenance mechanisms are a key feature of cancer cells and an attractive therapeutic target promising cancer specificity with few side effects. Cancer cells can either express the reverse transcriptase TERT or induce a recombinational process called Alternative lengthening of telomeres (ALT) (48). Much work has been done on targeting telomerase in cancers, yet no anti-telomerase therapeutic has excited clinical trial successfully (55). Furthermore, cell line models have also shown that targeting telomerase may lead to the emergence of ALT-positive clones (128, 129). In theory, this means that any successful treatment regimen for TMM must target both telomerase and ALT pathways. As no ALT-specific drug exists, there is a critical need for understanding mechanistic implications of ALT and what susceptibilities are introduced in these cancers. ALT presence is confirmed by cytidine-rich extrachromosomal telomere repeats (C-circles), exclusively found in ALT cancers (66). C-circles are byproducts of telomere recombination (8), and thus represent ALT-activity as well. Mutations within the repeat repressors ATRX/DAXX are overrepresented in ALT (84). While many studies have focused on the telomere-specific effects of this pathway, and indeed so has most other ALT-related research, it is vital to understand how the general chromatin environment cooperates in promoting ALT. In this thesis, I investigated three main areas of ALT research. (i) ALT-activity in sarcoma biopsies. (ii) Molecular features of ALT from different sequencing-based readouts and their integration into an omics signature. (iii) Inhibiting epigenetic marks in ALT cell lines using drugs.

To address these questions, I performed a C-circle screen of more than 600 mixed-lineage sarcoma samples. The resulting analysis showed that ALT-activity is heterogeneous. Next, by assembling a heterogeneous ALT cell line panel, I performed small and long RNA-seq together with ATAC-seq. ALT presence was found in chromatin accessibility, gene expression, and miRNA expression. Increased chromatin accessibility was driven by TFs within the AP-1 and RUNX family, whilst decreased accessibility was linked to SOX TFs. The upregulated miRNAs 138/142 could explain the lower activity of SOX4, whereas the downregulation of miR-218 could increase RUNX1 activity. Upregulated ALT genes had enrichment of immune-related TFs, which may be indicative of inflammation. GO-term analysis revealed that developmental genes were both up and downregulated, whereas pathway analysis further saw the presence of angiogenesis and activin/inhibin signaling. By integrating ATAC and transcriptome, the putative ALT biomarkers *BRSK2* and *NFATC2* could be found. Additionally, by performing multi-omics integration, an ALT signature could be extracted. Lastly, by testing different compounds in terms of cell viability, it was found that disrupting H3K27me3 and DNA methylation led to differences in viability relating to ALT activity. ALT cell lines with a high ALT activity were more sensitive to H3K27me3 inhibition but more resistant to DNA methylation inhibition.

In conclusion, this thesis characterized three areas that are vital for current and future ALT research. ALT cancers have a very heterogeneous ALT activity, which needs to be accounted for when identifying common features. Nevertheless, different omics readouts could be correlated with the presence of ALT in a cell line panel. ALT specific features included specific TF families as well as deregulated pathways. Additionally, integrating the different omics could be used to define a multi-omics ALT signature. Lastly, inhibiting repressive epigenetic modifiers revealed some correlation between cell viability and ALT activity.

Materials and Methods

1 Materials

1.1 Kits

Kit	Reference number	Company
Bioanalyzer high sensitivity DNA kit	5067-4626	Agilent Technologies, USA
CellTiter Glo	G7571	Promega, USA
Clarity western ECL blotting substrate	1705061	Bio-Rad, USA
MinElute PCR purification Kit	28004	QIAGEN, Germany
MODified histone peptide array	13005	Active-Motif, USA
NEBNext HF 2X PCR master mix	M0541S	New England Biolabs, USA
NEBNext multiplex oligos for Illumina	E7500S, E7780S	New England Biolabs, USA
NEBNext multiplex small RNA library prep set for Illumina	E7560, E7580	New England Biolabs, USA
NEBNext rRNA depletion kit	E6310X	New England Biolabs, USA
NEBNext ultra II RNA library prep kit for Illumina	E7770L	New England Biolabs, USA
phi29 DNA polymerase kit	M0269S	ThermoFisher Scientific, USA
QuBit dsDNA HS assay kit	Q32851	New England Biolabs, USA
SP NucleoSpin kit	740971.50	Macherey-Nagel, Germany
TeloTAGGG telomere length assay	12209136001	Merck, USA

1.2 Reagents and chemicals

Compound name	Reference number	Company
0.4 % tryphan blue	T13001	Logos Biosystems, South Korea
5-azacytidine	S1782-1ML	Selleckchem, USA
Agencourt AMPure XP	A63881	Beckman Coulter, USA
Agencourt RNACLEAN XP	A63987	Beckman Coulter, USA
AZD7648	S8843-5MG	Selleckchem, USA
BIBR-1532	S1186	Selleckchem, USA
BSA 20 mg/ml	B9000S	New England Biolabs, USA
CC1 pan-kinase inhibitor	N266A	Promega, USA
cGAMP	tlrl-nacga23-02	InvivoGen, USA
Colcemid in HBSS	D1925-20ML	Sigma-Aldrich, USA
CP2 cyclic peptide	S8601-1MG	Selleckchem, USA
CX-5461	18392-5MG	Cayman Chemicals, US

DAPI	62248	ThermoFisher Scientific, USA
DIG easy hyb granules	11796895001	Roche, Switzerland
DMSO	D2650-100ML	Sigma Aldrich, USA
dNTP mix	R0191	ThermoFisher Scientific, USA
Epredia EZ double cytofunnel	A78710005	ThermoFisher Scientific, USA
Falcon 70 µm cell strainer	352350	Corning, USA
G10	S8954-5MG	Selleckchem, USA
GSK343	2281-5MG	BioVision, USA
IKK-16	S2882	Selleckchem, USA
ImmEdge hydrophobic barrier PAP Pen	H-4000	Vector Laboratories, USA
KU-55933	S1092	Selleckchem, USA
LDN-192960	S3406-5MG	Selleckchem, USA
LUNA cell counting slides	L12001	Logos Biosystems, South Korea
LY2603618 (rabusertib)	S2626-5MG	Selleckchem, USA
MLN8237	S1133-5MG	Selleckchem, USA
NEBNext multiplex oligos for Illumina (dual index primers)	E7780S	New England Biolabs, USA
Phosphatase inhibitor cocktail	1862209	ThermoFisher Scientific, USA
Pierce 16% Formaldehyde methanol-free	28908	ThermoFisher Scientific, USA
Pierce 96-well polystyrene plates, white opaque	15042	ThermoFisher Scientific, USA
ProLong Diamond antifade mountant	P36965	ThermoFisher Scientific, USA
ProLong Gold antifade mountant with DAPI	P36931	ThermoFisher Scientific, USA
Propidium iodide	P3566	ThermoFisher Scientific, USA
QuBit assay tubes	Q32856	ThermoFisher Scientific, USA
Roti-Nylon plus 0.45 µM	K058.1	Carl Roth, Germany
RU.521	inh-ru521	InvivoGen, USA
Shandon Cytoslide	5991056	ThermoFisher Scientific, USA
Shandon filter cards	5991022	ThermoFisher Scientific, USA
SYBR green	S7536	ThermoFisher Scientific, USA
Tn5 tagmentase	15027865	Illumina, USA
Tn5 tagmentase buffer	14027866	Illumina, USA
ZM44739	S1103-5MG	Selleckchem, USA

1.3 Buffers

Buffer	**Composition**
Annexin V binding buffer (1X)	10 mM HEPES, 2.5 mM $CaCl_2$, 140 mM NaCl, dissolved in PBS (1x)
Hypotonic cell swelling buffer (KCl, 1X)	100 mM KCl dissolved in PBS (1x)
PBS (pH 7.4, 20X)	7 mM KCl, 1.7 mM KH_2PO_4, 137 mM NaCl 10 mM Na_2HPO_4

SSC (20X)	3 M NaCl, 0.3 M $C_6H_8O_7Na_3$. pH 7
KCM buffer	120 mM KCl, 20 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.1 % Triton

1.4 Antibodies

Antibody	Reference number	Company	Dilution
Anti-H1 (mouse)	05-457	Merck, USA	1:5000
Anti-H3.3S31p (rabbit)	39637	Active Motif, USA	1:100
Anti-CENP-A (mouse)	ADI-KAM-CC006-E	Enzo Life Science, USA	1:250
Goat anti-mouse alexa Fluor 488	A11029	ThermoFisher Scientific, USA	1:250
Goat anti-rabbit alexa Fluor 488	Ab150077	Abcam, UK	1:250
Goat anti-rabbit alexa Fluor 568	A11036	ThermoFisher Scientific, USA	1:250
Anti-mouse IgG HRP linked antibody	#7076	Cell Signaling, USA	1:2000

1.5 Cell culture reagents

Medium/supplement	Reference number	Company
Amniopan complete medium	P04-70100	PAN-biotech, Germany
Cryo-SFM	C-29910	Promocell, Germany
DMEM high glucose, pyruvate, no glutamine, no phenol red	31053028	ThermoFisher Scientific, USA
DMEM low glucose, pyruvate, no glutamine, no phenol red	11880036	ThermoFisher Scientific, USA
Fetal bovine serum (FBS) Standard	P30-3306	PAN-biotech, Germany
L-glutamine	P04-80100	PAN-biotech, Germany
Penicillin-Streptomycin	P06-07050	PAN-biotech, Germany
Trypsin	P10-023100	PAN-biotech, Germany

1.6 Primers

Primer name	Sequence
Intronic *POU5F1* forward primer	TTGGCTCTGGACCTTATCCC
Intronic *POU5F1* reverse primer	TTCCATCACTGGCTCGTAGC
ND4 forward primer	ACAAGCTCCATCTGCCTACG
ND4 reverse primer	TTATGAGAATGACTGCGCCG

1.7 Instruments

Instrument	Company
Andor Dragonfly 500	Oxford Instruments, UK
Bio-dot apparatus	Bio-Rad, USA
Chemidoc XRS+	Bio-Rad, USA
DMI6000 TCS SP5	Leica Microsystems, Germany
HiSeq 2000	Illumina, USA
HiSeq 4000	Illumina, USA
LSRFortessa II	Becton, Dickinson and Company, USA
Luna cell counter	Logos Biosystems, South Korea
NovaSeq 6000	Illumina, USA
Qubit 4.0	ThermoFisher Scientific, USA
Shandon Cytospin 2	ThermoFisher Scientific, USA
Shandon Cytospin 4	ThermoFisher Scientific, USA
StepOnePlus real-time PCR system	ThermoFisher Scientific, USA
T100 thermal cycler	Bio-Rad, USA
Tapestation 2200	Agilent Technologies, USA
Tecan Infinite M200	Tecan, Switzerland

1.8 Software

Software	Version	Reference
Array analysis software	16.1	(130)
Ataqv	1.1.1	(131)
Bamtools	2.5.1	(132)
BEDTools	2.29.2 (nf-core atacseq), 2.30.0 (system)	(133)
Bowtie1	1.2.3	(134)
BWA	0.7.17-r1188	(135)
Deeptools	3.4.3	(136)
FASTQC	0.11.9	(137)
FASTX	0.0.14	(138)
FeatureCounts	2.0.1	(139)
FlowJo	10.8.0	(140)
GraphPad	9.2.0	(141)
HOMER	4.9	(142)
IGV	2.9.4	(143)
ImageLab	6.1	(144)
JAVA	11.0.9.1	(145)
MACS2	2.2.7.1	(146)
MiRDeep2	2.0.1.2	(147)
Mirtop	1.2	(148)
mirtrace	1.9	(149)
MultiQC	1.9	(150)
NF-core	20.10.0	(151)

NF-core atacseq	1.2.1	(152)
NF-core smrnaseq (modified)	1.1 dev	(153)
Picard	2.23.1	(154)
Pysam	0.15.3	(155)
Python	3.8.2	(156)
R	3.6.2 (NF-core atacseq), 3.6.3 (NF-core smrnaseq), 4.0.2 (system)	(157)
RSEM	1.3.0	(158)
RSeQC	2.6.6	(159)
RStudio Server	1.4.1717	(160)
Samtools	1.9 (NF-core smrnaseq), 1.8 (RNA-seq pipeline)	(161)
SortMeRNA	2.1	(162)
STAR	2.5.3	(163)
Trim Galore!	0.6.4_dev	(164)

1.9 R-packages

Package	**Package version**	**Reference**
BiocManager	3.12	https://cran.r-project.org/web/packages/BiocManager/
ChIPseeker	1.26.0	(165)
DEseq2	1.30.1	(166)
EnhancedVolcano	1.8.0	(167)
GenomicRanges	1.42.0	(168)
ggplot2	3.3.3	(169)
ggpubr	0.4.0	(170)
ggrepel	0.9.1	(171)
Limma	3.46.0	(172)
magrittr	2.0.1	(173)
MOFA2	1.1.15	(174)
Msigdbr	7.2.1	(175)
Org.Hs.eg.db	3.12.0	(176)
PCAtools	3.13	(177)
RColorBrewer	1.1.2	(178)
reticulate	1.18.9007	(179)
Rsubread	2.4.2	(180)
rtracklayer	1.50.0	(181)
Tidyverse	1.3.0	(182)
TxDb.Hsapiens.UCSC.hg38.knownGene	3.10.0	(183)

1.10 External data sources

Database	Version	Reference
ENSEMBL gene annotation version 94	94	http://ftp.ensembl.org/pub/release-94/
ENSEMBL regulatory build for GRCh38	20210107	(184)
Human reference genome index GRCh38	GRCh38	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
Mirbase	22	ftp://mirbase.org/pub/mirbase/CURRENT/
MSigDB	V6.0.C5	https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/MSigDB_v6.0_Release_Notes
piRNAdb	1.7.6 homo sapiens	https://www.pirnadb.org/download/downloadarchive/pirna/piRNAdb.hsa.v1_7_6.fa.zip
Repeat masker	4.1.2	https://www.repeatmasker.org/RepeatMasker/RepeatMasker-4.1.2-p1.tar.gz
TelNet database	N/A	(185)

2 Cell culture

An overview of the cell line and growth conditions can be found in Table 2. Cal72, U2OS, Saos2 were purchased from DSMZ (Braunschweig, Germany). HOS and MG.63 were purchased from CLS Cell line services GMBH (Eppelheim, Germany). Pediatric glioblastoma cell lines were a kind gift from Prof. Stefan Pfister (DKFZ, Heidelberg, Germany). ATRX knockout cell lines (NEM168 ATRX KO clone B5 and F2) were generated previously in the lab by Dr. Inn Chung. Cell line identity was verified by single nucleotide polymorphism profiling (Multiplexion GmbH, Germany) and regularly tested for mycoplasma contamination with VenorGeM Advance. Cells were cultured at 37 °C with 5 % CO_2, split by washing flasks with PBS twice, and by dissociation with trypsin. Flasks were agitated every 2 minutes, and detachment was visually confirmed with a brightfield microscope. Trypsin was deactivated using 3x complete medium. Media were kept at 4 °C for a maximum of three months for DMEM based medium and two months for Amniopan.

Table 2. Cell line growth condition.

Tumor entity	Cell lines	Growth condition	Cryopreservation
Osteosarcoma	U2OS, Saos2, MG.63, Cal72, HOS	Low glucose DMEM, 10 % FBS, 2 mM L-glutamine, 100 µg/ml penicillin/streptomycin	20 % DMSO, 20 % FBS, 60 % medium
Pediatric glioblastoma	KNS42, MGBM1, SF188	High glucose DMEM, 10 % FBS, 2 mM L-glutamine, 100 µg/ml penicillin/streptomycin	20 % DMSO, 20 % FBS, 60 % medium
Pediatric glioblastoma	NEM157, NEM165, NEM168, NEM168 ATRX $^{-/-}$ clones B5 and F2	Amniopan complete medium	Cryo-SFM

3 C-circle screen of 688 mixed-lineage sarcomas

3.1 C-circle amplification of sarcoma DNA samples and analysis

Based on the initial protocol from Henson et al. (186), a quantitative C-circle assay was developed to analyze large sets of primary tumor samples. Genomic DNA was obtained from Dr. Priya Chudasama (NCT, Heidelberg, Germany), and C-circle assays were conducted with technical support from Caroline Knotz (Division of Chromatin Networks, DKFZ, Heidelberg, Germany). 688 mixed-lineage sarcoma DNA samples were obtained in H_2O and stored at -80 °C until amplification was performed. DNA content was first assayed using the QuBit 4.0 high-sensitivity program, and samples were diluted to contain 20 ng DNA per reaction. Six samples were prepared per membrane, where three included polymerase (+Pol) and three without polymerase (-Pol), as well as a U2OS standard curve for sample signal extrapolation. Each sample contained 5.05 µL nuclease-free H_2O, 0.1 M DTT, 2 µL phi29 buffer, 0.4 µL BSA, 0.2 µL 10 % Tween-20, and 0.8 µL dNTP mix. Additionally, samples with polymerase had 0.75 µL phi29 polymerase, whereas the -Pol controls contained the same volume in nuclease-free water. This was added to the 10 µL sample, briefly vortexed, spun, and placed in a thermal cycler. The reactions were incubated for 8 h at 30 °C and then at 20 min at 65 °C. A Roti Nylon plus membrane was used for dot-blotting and was hydrated prior using 2X SSC diluted in PBS. Samples were prepared by mixing 40 µl 2X SSC and then blotted onto the membrane using a Bio-Dot apparatus. The membrane was then heated at 120 °C for 20 min and hybridized using TeloTAGGG telomere length assay kit following the manufacturer’s protocol. Membranes were pre-hybridized with 18 mL DIG easy hyb for 60 min at 42 °C, and the solution was discarded. Meanwhile, 10 mL fresh DIG easy hyb was added together with 2 µL telomere probe. Hybridization was performed overnight at 42 °C, the membrane was washed in stringent wash buffer I, and incubated for 5 min at RT. The wash was done twice, and then the membranes were washed using stringent wash buffer II twice. These washes were done for 20 minutes at 50 °C. Membranes were then washed again using 1x SSC buffer for 5 min at RT and then blocked for 30 min at RT using the blocking solution provided in the kit. Anti-DIG-AP was diluted 1:10000 in blocking solution, added, and incubated with the membranes for 30 min at RT. The membranes were then rinsed twice for 15 min at RT using 1x SSC buffer, and detection buffer was added to the membranes. This was placed in the dark for 5 min at RT, then discarded. Membranes were visualized using chemiluminescence in the Chemidoc High-sensitivity program. 10-40 images were taken over the course of 300 seconds, and the image which contained the longest exposure without oversaturation was chosen. Oversaturated images were used for creating masks in ImageLabs, and densitometric values were extracted. Each +Pol sample was normalized to the sample that lacked polymerase and background. Then, the +Pol signal was averaged for each sample, and the C-circle ratio relative to the U2OS standard was calculated.

3.2 Data transformation and statistical testing

The C-circle ratios were log2 transformed using the equation $\mathrm{Y} = \mathrm{Log2}(X + 0.05)$ with X being the untransformed C-circle ratio, and Y being the final transformed C-circle ratio. Fitting of the distribution to a sum of two Gaussians was performed with Graphpad Prism using nonlinear regression (curve fit) function. Mean, standard deviation, and other statistical metrics were obtained from this analysis and used to calculate the 10th and 90th percentile for ALT- and ALT+ populations, respectively. This was done by using the equation $X = \mu + Z\sigma$, where µ stands for the population mean, Z for the Z-score and σ as the variance. X is defined in this case as the C-circle ratio where the nth percentile is contained within.

4 Cell viability assays

4.1 Luminescent cell viability assay

Cells were dissociated and counted using a Luna Cell counter. Clear flat bottomed 96 well plates were seeded for 24 h with 500 cells per well and 100 µL medium in total. Three wells per condition were reserved for medium only to account for background signal. Next, cells were treated for six days at 37 °C and 5 % CO_2. GSK343 was reapplied every third day due to heat instability. All treatment concentrations had equal amounts of DMSO, and the maximum DMSO concentration in the medium was below 0.2 %. Cell-Titer Glo, which measures cell viability by forming luminescence from ATP, was performed according to the manufacturer's recommendations. Plates were equilibrated at RT for 20 minutes; meanwhile, lysis reagent was prepared using a 1:1 solution together with complete medium and equilibrated at RT for 30 min. The supernatant in the plate was flicked off, and 200 µL of lysis:medium mixture was added per well. The plate was then agitated using a plate shaker at 600 RPM for 6 min. Wells were resuspended using a multichannel, with great care not to form bubbles. The samples were then transferred to an opaque white-walled 96-well plate and analyzed using a Tecan plate reader (luciferase program, integration time 1 ms). Each condition was averaged, and then the background signal was removed and normalized to DMSO controls for each cell line. EC_{50} values were calculated using the log inhibitor vs. normalized response function in Graphpad Prism.

4.2 Cell death flow cytometry assay

50,000 cells were seeded in a T-25 24 h prior and then treated with DMSO or 6 µL GSK343 for six days. GSK343 was reapplied every third day. The medium was collected together with trypsinized cells and spun at 300 x g for 5 min. 2 million cells were used per sample and resuspended in 1 mL Annexin binding buffer. Samples were kept on ice whenever possible and washed twice with ice-cold PBS. Samples were strained using a 70-micron cell strainer mesh and transferred to FACS tubes. 5 µL of Annexin V FITC and 10 µL of propidium iodide were added and incubated for 15 min RT in the dark. 400 µL Annexin binding buffer was added, and the samples were transported on ice and analyzed using a FORTESSA II in the DKFZ flow cytometry core facility. Each cell line contained single-stain controls for propidium iodide and Annexin V FITC for downstream compensation, and experiments without 50,000 detected events were discarded. FlowJo was used for gating and fluorescence compensation, and % live *versus* dead was calculated in Excel. Dead cells were defined as any event that was positive for positive propidium iodide/FITC or both. Live cells were defined as propidium iodide and FITC negative.

5 Analysis of histone H3 phosphorylation at serine 31 (H3.3S31p) during metaphase

5.1 H3.3S31p antibody validation with histone peptide array

To validate the specificity of the primary H3.3S31p antibody that was used in metaphase spreads (see section 5.2), the MODified histone peptide array was used. This array contains 384 combinations of histone modifications. This assays how combinations of histone modifications can impact primary antibody binding. It should be noted that the H3.3S31p peptide was not part of the histone peptide array. 3 ml blocking solution (provided in the kit) was added to the array and incubated on a shaker for 1 hr at RT. The blocking solution was discarded, and the array was rinsed in TTBS buffer (provided in the kit). The primary H3.3S31p antibody was added to the blocking solution and incubated with the array for 1 hr at RT. Following this, the array was quickly rinsed with TTBS buffer and then washed three times for 5 min with new TTBS buffer. The secondary HRP conjugated antibody was added to the blocking solution and incubated with the array for 1 hr RT in the dark. The array was then washed again, three times for 5 mins with TTBS buffer, and Clarity ECL substrate was prepared by mixing Clarity western peroxide reagent and Clarity western luminol reagent 1:1. 5 ml of the ECL was added onto the array, incubated in the dark for 5 min RT, and then the chemiluminescence was detected in the Chemidoc XP.

The resulting detection can be found in Figure 5. The non-specific binding was then assayed using the Array analyze software provided from Active-motif. The resulting analysis is found in Figure 6. Without H3.3S31p on the array, it is challenging to know the ratio between H3.3S31p and non-specific binding. However, the non-specific signal is still readily detected through chemiluminescence. The main hit was H4K12ac, followed by H4R19me2 and H4K5ac. Histone 4 (H4) marks are not as well characterized as the Histone 3 counterparts; however, both H4K12ac and H4K5ac are also centromeric marks (187). Future studies with recombinant H3.3S31p will be informative in observing how much of the centromeric H3.3S31p signal consists of these H4 marks.

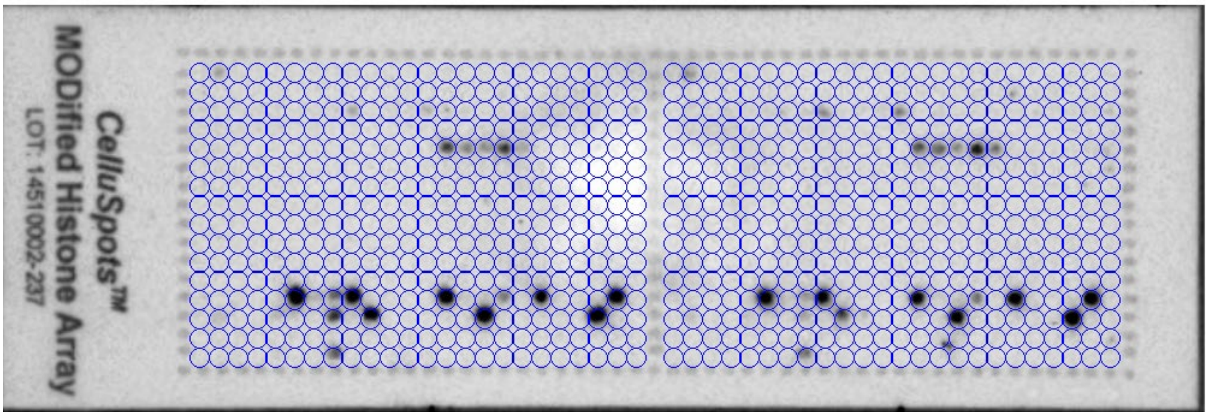


Figure 5. Chemiluminescent detection in the histone peptide array. 1:100 primary H3.3S31p was used, and the non-specific binding to other histone peptides was imaged

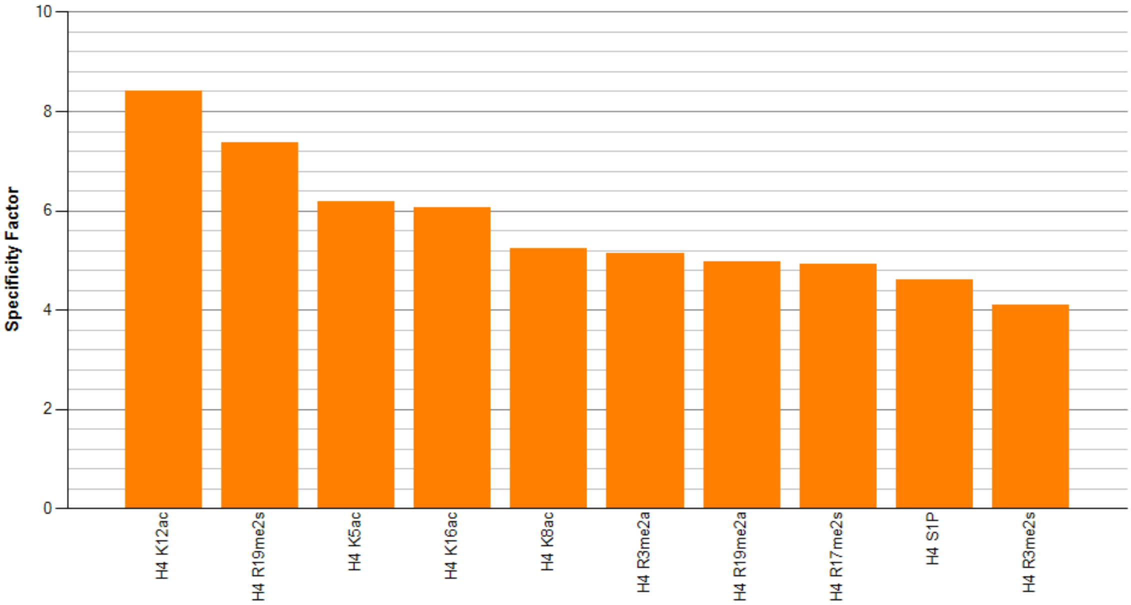


Figure 6. Output from Array analysis software regarding H3.3S31p non-specific binding. The specificity factor is calculated by dividing the signal intensity for the histone peptide by the background staining.

5.2 Optimized metaphase spreads of mitotically arrested cells

Metaphase spreads were performed with technical assistance from Caroline Knotz. 200,000 MGBM1 cells were seeded in a 10 cm dish 24 h prior to exposure. 12.5 µL Colcemid was added to each plate, together with the respective inhibitor or equivalent volume of DMSO. 16 h later, the cells were dissociated using trypsin and centrifuged at 500 x g for 10 min. The supernatant was removed, and pellets were washed with 400 µL PBS. 10 µL cell pellet was mixed with 10 µL tryphan blue, and cell number and viability were assessed using the Luna cell counter. 30,000 cells were used for metaphase spreads and were resuspended in 1 mL KCl hypotonic buffer for 10 min at RT. Cytospin funnels and slides were prepared in parallel. Slides had been previously cleaned with 80 % ethanol and stored at – 20 °C until use. EZ double funnels and Shandon filter papers were mounted onto the slides, and 500 µL sample was added per funnel. Cells were spun on a Shandon Cytospin 4 at 1,600 RPM for 10 min, and cell spots were demarcated using a hydrophobic barrier pen. Great care was taken as not to dry the slides, and they were immediately fixed using 4 % PFA in PBS for 10 min at RT. The slides were washed three times with PBS and then permeabilized using 0.02 % TX-100 in PBS for 5 min. Following a brief PBS wash, the slides were blocked using 4 % BSA in PBS for 1 hr at RT, meanwhile having been placed in a humidity chamber. The slides were then washed three times for 5 min with PBS, and 50 µL primary antibody diluted in blocking solution was added onto each cell spot. Slides were placed in a humidity chamber for 1 hr at RT and then washed three times for 5 min with PBS each. Secondary antibodies were added, and the slides were washed in the same manner. However, the slides were protected from light. DAPI was diluted in PBS 1:2,000, and 50 µL was added onto each cell spot for 15 min at RT and in a humidity chamber. Following a triple rinse with PBS, the cells were rinsed with water and then 70 % ethanol. Dehydration was performed by submerging the slides in 100 % ethanol for 2 min and then air-dried for 5 min. Slides were mounted using ProLong Diamond and coverslips and allowed to cure for 24 h. Lastly, the coverslips were sealed using nail polish and imaged using either the Leica SP5 (60x oil objective) or Andor Dragonfly (100x silicone objective). Chromosome segmentation was done using FIJI, where images were pre-processed using Gaussian blurring. Segmentation was then performed on the DAPI channel using Otsu's method and overlayed with the H3.3S31p channel. Intensities were extracted using the ROI manager.

5.3 Metaphase spreads of mitotically arrested cells exposed to different inhibitors

Confluent MGBM1 cells in a T-25 were split 1:3 and grown until 70 % confluent. At this stage, the cell line was treated with 5 µL Colcemid, together with inhibitors or DMSO. After 4 hours, the flasks were vigorously shaken to detach mitotic cells, and the medium was collected and spun at 300 x g for 5 mins. Cells were then counted, and 2,000 cells were resuspended in 1 ml 0.5 % sodium citrate in H_2O. The sample was incubated for 10 min at 37

°C and 5 % CO_2 and then put on ice. The samples were on ice for approximately 15 min and then loaded onto a Cytospin 2 with a funnel sandwich. The funnel sandwich consisted of EZ double funnels together with filter paper and cytospin slides. 500 µL sample was loaded onto each funnel and spun at 1,500 revolutions per minute (RPM) for 10 min. The slides were dried and then put in KCM buffer for 15 min RT. The cell spots were then demarcated using a hydrophobic barrier pen and fixed with 4 % PFA in PBS for 10 min in RT and washed two times in KCM buffer for 5 min each. Slides were then incubated with primary H3.3S31p antibody dissolved in 10 % goat serum in PBS overnight in a humidity chamber. Following this, the slides were washed two times in KCM buffer for 5 min. Secondary alexa fluor 488 was diluted in 10 % goat serum in PBS and added onto the slides for 1 hr RT in a humidity chamber. The cells were then washed for 5 min in H_2O and dried. Lastly, slides were mounted with ProLong Gold with DAPI and a coverslip and allowed to cure for 24 hr in RT and protected from light. Image acquisition was made with a Leica SP5.

6 RNA-seq of long and small RNA and subsequent computational methods

6.1 Sample lysis and preparation for long and small RNA-seq

A total of $1 \cdot 10^6$ cells were processed and lysed according to the manufacturer's recommendation using the MinElute kit, with technical assistance from Caroline Knotz. Briefly, 300 µl ML buffer was added, and the cell pellet was resuspended by pipetting. Following 5 min incubation at RT, the lysis mixture was added onto a NucleoSpin filter and centrifuged for 1 min at 11,000 x g. The filter was discarded, and 100% ethanol was added to the flow-through. Samples were briefly vortexed and incubated for 5 min at RT, then placed onto a NucleoSpin RNA column. The flow-through, containing small RNAs > 200 nt, was saved. The column containing long RNAs < 200 nt was then subjected to desalting by adding 350 µl MDB buffer and centrifuged for 1 min at 11,000 x g.

The saved fraction containing small RNAs underwent protein precipitation by adding 350 µl MP buffer, vortexed, and then centrifuged for 3 min at 11,000 x g. The supernatant was carefully removed by pipetting and then placed on a NucleoSpin protein removal column. Samples were centrifuged for 1 min at 11,000 x g, columns were discarded, and the flow-through was mixed with 800 µl MX buffer. 725 µl mixture was added onto a new NucleoSpin

RNA column, centrifuged for 30 s at 11,000 x g. The flow-through was discarded, and the remaining sample mixture was loaded on the same column and centrifuged as previously. 600 µl MW1 buffer was added and then centrifuged for 30 s at 11,000 x g, the flow-through was discarded, and then 700 µl MW2 buffer was added to the columns and spun as previously. 250 µl MW2 buffer was added and centrifuged for 2 min at 11,000 x g, ensuring that the column's silica membrane was dry. The small RNA bound in the column was then eluted by adding 30 µl nuclease-free H_2O.

6.2 Small RNA-seq library preparation

The small RNA was prepared using the NEBNext multiplex small RNA library prep set. 600 ng RNA was used, and nuclease-free H_2O was added so that each sample contained 6 µL. 1 µL 3’ SR adaptor for Illumina was added and incubated in a thermal cycler for 2 min at 70 °C. Then, 10 µL 3’ ligation reaction buffer and 3 µL 3’ ligation enzyme mix were added. Following a 1 hr incubation at 25 °C, 1 µL SR RT Primer for Illumina and 4.5 µL nuclease-free water were added. Samples were then placed in a thermal cycler at the following program: 1) 5 min at 75 °C. 2) 15 min at 37 °C. 3) 15 min at 25 °C. In parallel, 5’ SR adaptor was diluted in 120 µL nuclease-free water and then denatured in a thermal cycler for 2 min at 70 °C. 1 µL 5’ SR adaptor was added to the samples, together with 1 µL 5’ ligation buffer and 2.5 µL 5’ ligation enzyme mix. The samples were placed anew in a thermal cycler for 1 hr at 25 °C, and then the reverse transcription was performed by adding 8 µL first-strand synthesis reaction, 1 µL murine RNase inhibitor, and 1 µL ProtoScript II reverse transcriptase. Samples were again placed in a thermal cycler for 1 hr at 50 °C. Following this, PCR amplification was done by adding 5 µL nuclease-free water, 2.5 µL index primer, 2.5 µL SR primer for Illumina, and 50 µL LongAmp Taq master mix. The samples were mixed well and then placed in a thermal cycler at the following program: 1) 30 sec at 94 °C. 2) 12 cycles of 15 sec at 94 °C, 30 sec at 62 °C, 15 sec at 70 °C. 3) 5 min at 70 °C. The samples were then cleaned up using AmPure XP beads, where 180 µL beads were added per sample. Following a 5 min incubation at RT, the tubes were placed on a magnetic rack for 2 min at RT for separation. Supernatant was discarded, and 200 µL 80 % ethanol was added gently as not to disturb the beads. This was incubated for 30+ sec at RT, gently removed, and then the ethanol wash was repeated once again. The beads were left to dry for 10 min at RT, and cDNA was eluted using 25 µL nuclease-free water. The tubes were placed on a magnetic rack for 2 min at RT, and the supernatant was gently pipetted into a new tube. cDNA content was assessed using QuBit high sensitivity program, and a 0.5 µL sample was used to determine library size using the Tapestation high sensitivity screentape. Samples were submitted for 50 bp single-read sequencing at the DKFZ genomics core facility on the HIseq 2000.

6.3 Long RNA-seq library preparation

The long RNAs from above were depleted for rRNA using the NEBNext rRNA depletion kit. First, libraries were hybridized with probes by adding 11 µl total RNA, 2 µl NEXTNext v2 rRNA depletion solution, and 2 µl NEBNext probe hybridization buffer for a total volume of 15 µl per reaction. Samples were then mixed thoroughly by resuspension, briefly spun, and then placed in a thermocycler with the following program: 2 min at 95 °C, cooling down to 22 °C (0.1°C /sec) and then held at 22 °C for 5 min. Samples were then spun and placed on ice. The samples were then RNase H digested on ice by adding 2 µl RNase H reaction buffer, 2 µl NEBNext thermostable RNase H, and 1 µl nuclease-free water for a total volume of 20 µl. The samples were again mixed thoroughly by resuspension, briefly spun, and then placed in a thermocycler for 30 min at 50°C. The samples were then subjected to DNase I digestion by adding 5 µl DNase I reaction buffer, 2.5 µl NEBNext DNase I, and 22.5 µl nuclease-free water for a total volume of 50 µl. The samples were placed in a thermocycler for 30 min at 37 °C, and then the RNA was purified using Agencourt RNAClean XP beads. 90 µl 1.8X beads was added, mixed thoroughly by resuspension, and then incubated for 15 min on ice. Then the tube was placed on a magnetic rack, where the supernatant was carefully removed and discarded. 200 µl 80 % ethanol was added to the beads, incubated at RT for 30 s, and then the supernatant was carefully discarded. The whole bead purification was repeated twice, and then residual ethanol was removed by air drying the beads for up to 5 min. The tubes were then removed from the magnetic rack, and the RNA was eluted by adding 7 µl nuclease-free water. The samples were mixed thoroughly and then incubated at RT for 2 min. Then the beads were separated by putting the tubes back on the magnetic rack. The supernatant, now containing the long RNA, was transferred to a new tube and stored at -80 °C. The RNA was then subjected to library preparation, using NEBNext ultra II according to the manufacturer’s recommendation. The RNA was firstly assessed using High sensitivity RNA screentape, where 0.5 ng RNA was added per lane. Samples containing RNA integrity number (RIN) below 7 were discarded; otherwise, the samples were fragmented and primed by adding 5 µl sample, 4 µl

NEBNext first-strand synthesis reaction buffer, and 1 µl random primers for a total volume of 10 µl. This mixture was mixed by resuspension and placed on a thermocycler for 7 min at 94 °C. The tubes were then put on ice, and cDNA was synthesized by adding 8 µl nuclease-free water and 2 µl NEBNext first-strand synthesis enzyme mix for a total volume of 20 µl. The sample was then put on a thermocycler with the following program: 10 min at 25 °C, 15 min at 42 °C, 15 min at 70 °C, and then held at 4 °C. 8 µl NEBNext second-strand synthesis reaction buffer, 2 µl NEBNext second strand synthesis enzyme mix and 48 µl nuclease-free water was added to a total volume of 80 µl. The sample was resuspended and incubated for 1 hr at 16 °C and then purified using AmPure XP beads. 144 µl beads were added to the sample, mixed well by resuspension, and then incubated for 5 min at RT. Samples were spun briefly, then placed on a magnetic rack for bead separation. The supernatant was discarded, and 200 µl 80 % ethanol was added and then incubated for 30 s at RT. The supernatant was discarded, and the beads were again washed with ethanol twice. The beads were then air-dried for 5 min on the magnetic rack, and the tubes were removed from the magnetic rack. DNA was eluted by adding 53 µl 0.1X TE buffer, incubated for 2 min at RT, and placed on the magnetic rack. 50 µl supernatant was transferred to a clean PCR tube, and finally, the library was prepped using the supernatant and adding 7 µl NEBNext Ultra II end prep reaction buffer and 3 µl NEBNext Ultra II end prep enzyme mix for a total volume of 60 µl. Samples were resuspended and placed in a thermocycler with the following program: 30 min at 20 °C and 30 min at 65 °C. Adaptors were then ligated by adding 2.5 µl adaptor, 1 µl NEBNext ligation enhancer, and 30 µl NEBNext ultra II ligation master mix for a total volume of 93.5 µl. Samples were resuspended at incubated for 15 min at 20 °C, then 3 µl USER enzyme was added. The library was again purified with AmPure XP beads, where 87 µl beads were added to the samples, resuspended, and then incubated for 5 min at RT. The tubes were then spun and placed on a magnetic rack for 5 min. The supernatant was discarded, and 200 µl 80 % ethanol was added, incubated at RT for 30 s, and then the supernatant was discarded again. This was repeated twice, and then the beads were air-dried for 5 min while on a magnetic rack. The DNA was eluted using 17 µl 0.1x TE buffer and incubated for 2 min at RT. The samples were then placed on a magnetic rack, and 15 µl supernatant was transferred to a new PCR tube. The DNA was then amplified using PCR, where 25 µl NEBNext ultra II Q5 master mix and 19 µl index prime mix were added for a total volume of 50 µl. The samples were placed on a thermocycler using the following program: 1) 30 sec at 94 °C. 2) 12 cycles of; 15 sec at 94 °C, 30 sec at 62 °C, 15 sec at 70 °C. 3) 5 min at 70 °C. Samples were submitted for 50 bp single-read sequencing to the DKFZ genomics core facility and analyzed using a HiSeq 4000.

6.4 Transcriptome alignment pipeline

Merged FASTQ files were aligned using a previously established in-house pipeline. FASTQ files were unzipped using gunzip, aligned to GRCh38 using STAR, and ribosomal reads were removed using SortMeRNA. Quality control was done using RSEM, RseQC, FastQC, and MultiQC. The resulting BAM files were utilized by featureCounts for count table generation, and genes were annotated using ENSEMBL (v.94).

6.5 Eigenvalues of correlation between principal components and cell line metadata for individual omics

Count data from each omics was variance-stabilizing transformed and subjected to principal component analysis (PCA) using the R-package PCAtools. The number of principal components was determined by the elbow method, and the Pearson correlation coefficient between cell line metadata and principal component was calculated using the function $eigencorplot()$. It was shown that most principal components had variations related to tumor entity, which was used then used for normalization within DEseq2 and multifactor omics analysis.

6.6 Transcriptome differential expression analysis

Count tables from ENSEMBL annotated genes were obtained from the pipeline and then used in DESeq2. Fold change cutoffs were set at 1.5, with a statistical cutoff of a Benjamini-Hochberg adjusted p-value of 0.01. Additionally, reads containing fewer than 5 normalized reads in 4 samples were discarded. Confounding effects of tumor entity were normalized by using the design factor: $design = \sim tumor\ entity + telomere\ maintenance\ mechanism$. Volcano plots were generated using EnhancedVolcano.

6.7 Gene-ontology of differentially expressed genes

Down and upregulated genes were separated and used in gene-ontology analysis (http://geneontology.org/, 2021-09-01 release). The category “biological process” was used, and the top 10 GO-terms with respect to the false discovery rate (FDR) were extracted.

6.8 Transcription factor motif analysis of promoters of differentially expressed genes

Differentially expressed genes were divided into down and upregulated, and the ENSEMBL IDs were then pasted into PASTAA (188), using ranked analysis and sorting according to the fold change of the genes. Motifs without known proteins were removed, and results were plotted using ggplot2 and ggrepel. Association score consists of the -Log of the most significant hypergeometric *p*-value.

6.9 TelNet annotation of differentially expressed genes

All differentially expressed genes were analyzed in the TelNet database (https://malone2.bioquant.uni-heidelberg.de/fmi/webd/TelNet) using the “list search” function. Gene name and functions were extracted, and all transcription factors were manually annotated by searching genecards (https://www.genecards.org/) for the terms “transcription factor”.

6.10 Normalized gene counts visualization

Gene counts were constructed first by collapsing replicate values using DESeq2 $collapseReplicate()$ function, and normalized counts were plotted using DEseq2 $\mathrm{plotCounts}(\mathrm{normalized} = \mathrm{True})$.

6.11 Pathway analysis with Reactome

Raw counts of all ensemble-annotated genes were averaged across replicates using R. These were then fed into Reactome (https://reactome.org) using the “analyze gene expression” function. Pathway analysis with down-weighting of overlapping genes (PADOG) was used for gene set enrichment analysis. Samples were compared in terms of ALT+/ALT- and the top 10 results in terms of FDR were used.

6.12 miRNA alignment pipeline

FASTQ files from small and long RNA-seq were merged for each replicate. These FASTQ files were processed using a newly established NF-core pipeline found here (https://nf-co.re/smrnaseq). The pipeline was forked from GitHub, and additional scripting was done by Simon Steiger (Division of Chromatin Networks, DKFZ, Heidelberg, Germany). The modified pipeline can be found here: https://github.com/Simontuk/smrnaseq/tree/v1-dev-alt.

Adapters were trimmed with Trim Galore!, and aligned to mature and hairpin miRNAs using bowtie1 to the Mirbase annotation (v 22) as a reference. Post-alignment processing was done using SAMtools and mirtop. Quality control was done with FastQC, MultiQC, edgeR, and mirtrace. Count tables were constructed by MiRDeep2

6.13 miRNA differential expression

Count tables from mature miRNAs were obtained from the pipeline and then used in DESeq2. Cutoffs were established as mentioned above, apart from not removing low count reads. DESeq2 analysis was done using the same design factor as above. Volcano plots were generated using EnhancedVolcano. miRNAs were additionally subgrouped using miEAA 2.0 (189) by choosing “annotations derived over miRTarBase (Gene Ontology)” where miRNAs with the gene ontology term “negative regulation of telomere maintenance via telomere lengthening” were extracted.

6.14 Transposable element and piRNA alignment pipeline

Transposable elements (TE) and piRNA were aligned using the in-house RNA-seq pipeline, except for permissive multimapping ($outFilterMultimapNmax = 100$) in STAR. TE was aligned with the repeat masker (4.1.2) database, and reads from individual TE transcripts were grouped into TE families. piRNAs were aligned with piRNAdb (v.1.7.6). In both instances, count tables were constructed using featureCounts and allowing for multimappers using $countMultiMappingReads = TRUE$.

6.15 Transposable element and piRNA differential expression

TE and piRNA count tables were used in DEseq2 with fold change and statistical cutoffs as previously mentioned, apart from not removing low count reads. Volcano plots were made with EnhancedVolcano.

7 ATAC-seq and subsequent computational methods

7.1 ATAC-seq tagmentation and quality control

ATAC-seq tagmentation and quality control were performed with technical assistance from Caroline Knotz. A total of $1 \cdot 10^5$ cells were harvested and washed with 1x ice-cold PBS and then spun at 300 x g for 5 h at 4 °C. The supernatant was discarded, and the cells were lysed with the following tagmented using the following recipe: 9.75 µL H_2O, 23.50 µL transposase buffer, 0.25 µL PIC, 2 µL Tn5 tagmentase, and 0.5 µL 0.5% digitonin. Samples were left to lyse for 30 min at 37 °C and then put on ice. Afterward, purification was done using Qiagen MinElute kit. 62.5 µl Buffer PB was added to the tagmented sample and placed onto a MinElute column. The sample was centrifuged for 1 min at 11 000 x g, and the flow-through was discarded. 750 µl PE buffer was added onto the column, spun as previously, and then the column was placed in a new collection tube. Columns were dried by spinning as previously and then placed into a 1.5 ml microcentrifuge tube. Each column was eluted in 12 µL EB buffer, where the flow-through was put onto the column one additional time. 2 µL of the sample was reserved for qPCR quality control. To this end, samples were diluted to 0.5 ng/µL in water. In total, six reactions were made from each sample. Three reactions were used as replicates for mitochondrial enrichment (*ND4*), and the other three were used as replicates for non-specific enrichment (intronic *POU5F1*). Dr. Lara Klett (Division of Chromatin Networks, DKFZ, Heidelberg, Germany) had previously validated these primers as reasonable proxies for ATAC quality. Each replicate contained 1 µL sample, 2 µL H_2O, 0.5 µL forward and reverse primer respectively, 1 µL 10x SYBR green and 5 µL NEBNext HF PCR master mix. These were mixed onto a 96 well plate and analyzed using the StepOnePlus Real-Time PCR system in the following program: 1) 5 min at 72 °C. 2) 30 sec at 98 °C. 3) 25 cycles of; 10 sec at 98 °C, 30 sec at 63 °C, 1 min at 72 °C. Values from qPCR were submitted to an Excel template made by me. The template averaged the mean Ct of the three replicates calculated the ΔCt enrichment calculation. Samples containing ΔCt enrichment of > 2.5 for mitochondrial reads and > 2 for non-specific products were discarded.

7.2 ATAC library preparation

The samples that passed the qPCR quality control were then mixed with 10 µL H_2O, 2.5 µL index primer 1, 2.5 µL index primer 2, 25 µL NEBNext HF PCR master mix, and 10 µL tagmented sample. PCR enrichment was done with the following program: 1) 5 min at 72 °C. 2) 30 sec at 98 °C. 3) 12 cycles of; 10 sec at 98 °C, 30 sec at 63 °C, 1 min at 72 °C. 4) 1 min at 72 °C. The samples were purified using Agencourt AMPure XP beads, where 80 µl beads were used per 50 µL reaction. The samples were incubated with the beads for 5 min at RT, then separated using magnets for 5 min. The supernatant was removed, and the pellet was washed twice with 80 % ethanol for five seconds per wash. The beads were dried (around 3-6 minutes) and eluted in 50 µL EB buffer for 5 min. Supernatant was used for additional purification using beads, where 25 µL beads were added per sample and incubated for 5 min at RT. The mixture was separated using magnets for 5 min, and the supernatant was transferred into new tubes. 45 µL beads were added anew, incubated for 5 min, separated with magnets for 4 min, and then washed twice with 80 % ethanol for 5 sec per wash. Samples were dried for 3-6 minutes and eluted in 15 µL EB buffer for 5 min. The DNA was measured with QuBit high sensitivity program, and 0.5 ng was added onto a Tapestation high-sensitivity DNA Screen Tape for library size analysis. The samples were then sequenced for 50 bp paired-end sequencing in the DKFZ genomics core facility using a NovaSeq 6000.

7.3 ATAC alignment pipeline

ATAC reads were processed using an atacseq NF-core pipeline (https://nf-co.re/atacseq/1.2.1), which was implemented with the help of Simon Steiger. Briefly, adapters were trimmed with TrimGalore!, and aligned with BWA to GRCh38. Duplicates were marked with picard and BAMTools. SAMtools. BEDtools and pysam were used for removing reads from blacklisted regions and other low-quality metrics. BigWig files were created using BEDTools. Additional quality control of the processed reads was done with ATAQV, FastQC, and MultiQC. Consensus peak sets and individual sample count tables were constructed with featureCounts.

7.4 ATAC differential peak analysis

Count tables of all samples were constructed by merging individual count tables generated by the pipeline. The Y-chromosome and decoy sequences were removed and then used for DESeq2. Fold change cutoffs were set at 1.5, with a statistical cutoff of a Benjamini-Hochberg adjusted p-value of 0.01. Additionally, reads containing fewer than 5 normalized reads in 4 samples were discarded. Confounding effects of tumor entity were normalized by using the design factor:

$design = \sim tumor\ entity\ +\ telomere\ maintenance\ mechanism.$ Volcano plots were generated using EnhancedVolcano.

7.5 ATAC differential peak annotation

Differential ATAC peaks were imported into R using Granges $\mathrm{readPeakFile}()$, and annotated using the ChIPSeeker R-package with the function $\mathrm{annotatePeak}()$. Pie diagrams were constructed using the $plotAnnoPie()$ function.

7.6 ATAC transcription factor motif analysis

Downregulated and upregulated peaks were separated into separate bed files using R, and transcription factor motifs were extracted using HOMER $annotatePeaks.pl$. Enrichment was calculated by dividing the % of motifs in the background *versus* the % of motifs in target sequences. Results were plotted using ggplot2 and ggrepel.

8 Omics integration

8.1 Integration of differential ALT+ ATAC peaks and expressed genes

The differentially expressed genes were converted into bed-format using ENSEMBL BioMart (https://m.ensembl.org/biomart/martview/). ATAC peaks were divided into upregulated and downregulated peaks, and promoter regions were extracted using ChIPSeeker and the function $getPromoters()$, with ENSEMBL regulatory build providing the promoter reference. Each promoter was annotated to a gene using the same reference. Enhancers were annotated by using the Bedtools $intersect\ -wa$ function and obtaining differentially accessible regions that overlapped with annotated enhancers from the ENSEMBL regulatory build. These were, in turn, annotated to genes, where ATAC-enhancer peaks inside gene bodies were annotated to respective genes, and ATAC-enhancer peaks outside gene bodies were annotated to the nearest gene. Gene names from enhancer/promoter regions and differentially expressed genes were extracted, as were all the genes in the TelNet database, and overlaps were found using http://www.interactivenn.net/.

8.2 ATAC RNA overlaps count and gene track visualization

Gene counts were constructed first by collapsing replicate values using DESeq2 $collapseReplicate()$ function, and normalized counts were plotted using DEseq2 $\mathrm{plotCounts}(\mathrm{normalized} = \mathrm{True})$. Gene tracks were made using bigwig files for ATAC-seq, and .tdf files for RNA-seq provided from the respective pipelines and imported into IGV.

8.3 Multi-omics factor analysis

Count tables for TE, piRNA, miRNA, ATAC peaks and expressed genes were first transformed using the DEseq2 function $assay(vst)$. These variance stabilizing transformed count tables were then normalized to tumor entity using Limma $RemoveBatchEffect()$.

The count tables were then imported into multifactor omics analysis (MOFA). Each count table was sorted in descending variance for each feature, and the top 100-10,000 features were used for comparison. 1,000 features were chosen for downstream analysis where the distinction between ALT+ and ALT- cell lines was the highest. The number of analyzed factors was determined by the elbow method. GO-term analysis of weighted genes was done using the MSigDB V6.0.C5 database and the MOFA function $run\_enrichment()$.

Results

1 C-circle screen of 688 soft-tissue sarcoma samples

1.1 C-circle screen workflow and representative image

C-circles are routinely used to assay alternative lengthening of telomeres (ALT) activity and have proven to be a promising biomarker for ALT-positive cell lines and primary cancers. As part of a collaboration with Priya Chudasama (NCT, Heidelberg, Germany), a quantitative C-circle assay was developed (Figure 7) to analyze 688 soft tissue sarcomas. This represents the most extensive set of patient samples of a tumor entity for which a C-circle assay normalized to a reference signal (U2OS) has been conducted. By using such a vast collection of primary tumor samples, I could describe the distribution of ALT-activity in primary cancer cells and assess ALT heterogeneity. By applying statistical modeling, the criteria for ALT+ samples could be more reliably defined than in previous studies. Figure 8 shows a representative image of a C-circle dot blot membrane, with ALT+ samples in red, and the - Pol controls shown with the high-exposure image. Each sample was performed in triplicates, where the + Pol denotes samples that were amplified by polymerase.

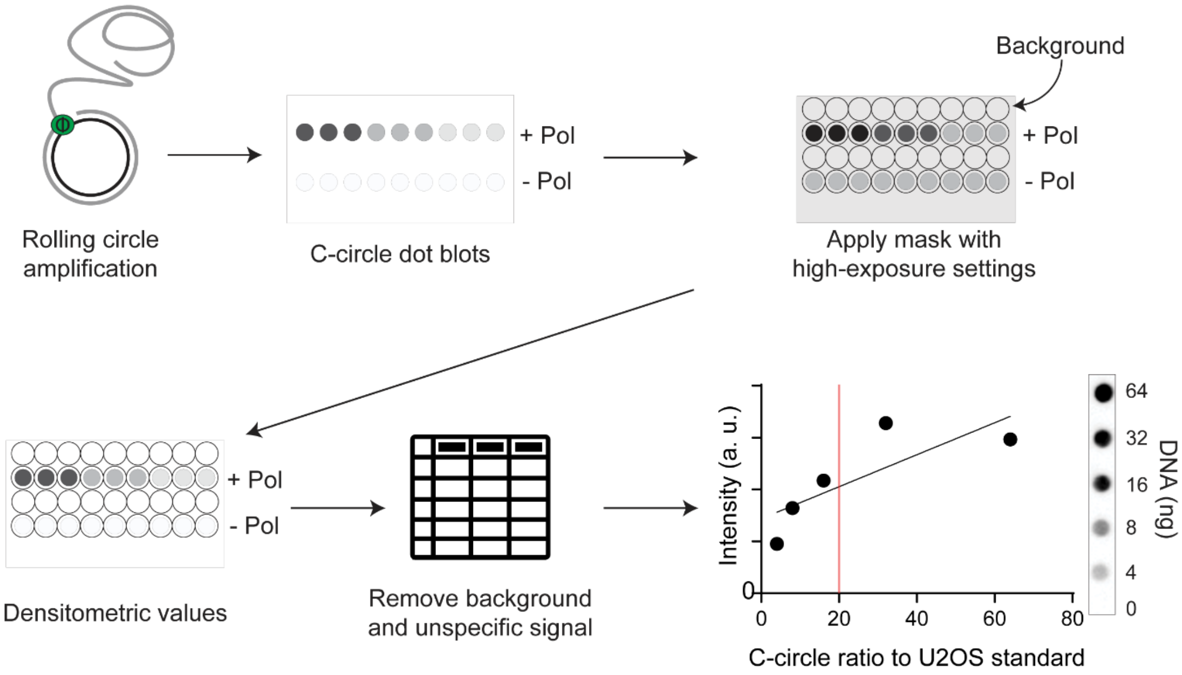


Figure 7. Workflow of C-circle screening of 688 mixed-lineage sarcomas.

To account for non-amplified circular DNA background, reactions omitting polymerase (- Pol) were also included. Telomere probes that allow for chemiluminescent detection were used. Each membrane was imaged in a time series through 300 sec, where the highest exposure without oversaturation was used for densitometry analysis. High-exposure conditions were used to aid in creating masks for dot localization and subsequent quantification. These masks encompassed + Pol and - Pol samples, as well as the background. Each + Pol reaction was then normalized to background as well as - Pol. The triplicate values were then merged into an average signal and extrapolated onto a standard curve. The C-circle signal was then expressed as a fraction of the U2OS signal using a standard curve. If the U2OS standard curves had R^2 values below 0.5, a standard curve with R^2 values above 0.5 was obtained from membranes performed on the same day. This quantitative C-circle approach yielded very robust results and set the threshold for ALT+ in the C-circle screen in a well-defined manner.

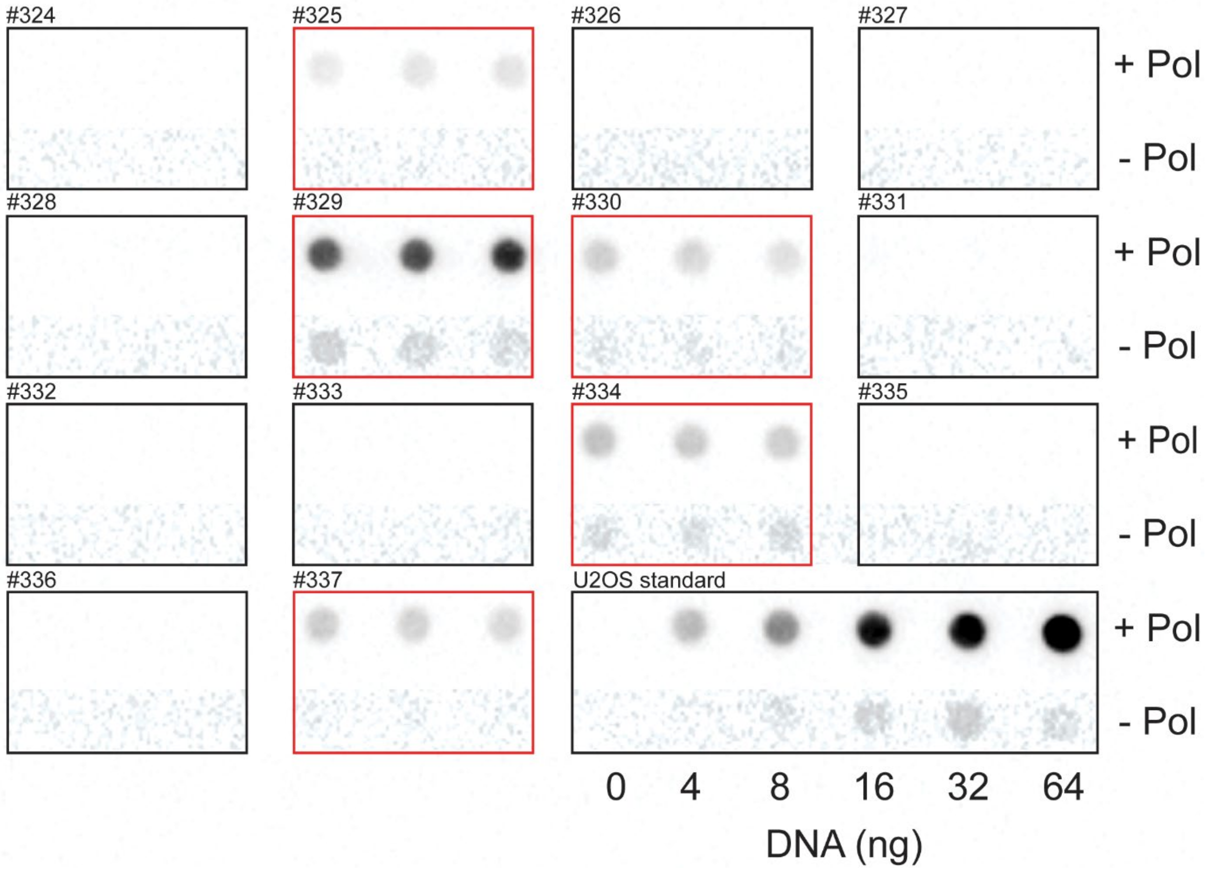


Figure 8. Representative image of C-circle membrane. + Pol denotes samples with complete C-circle amplification, and – Pol is the reaction background control. Each sample was performed in triplicates, and the U2OS standard curve was constructed using 6 different DNA concentrations. The – Pol is shown with high-contrast to show the reaction background of each sample.

1.2 The C-circle distribution shows that most samples have low detectable C-circles

The 687 samples that were successfully analyzed are plotted in Figure 9. Only one sample was omitted due to insufficient gDNA to repeat the C-circle amplification following an irregular standard curve. The scatter plot shows that most samples could be classified as having low C-circle ratios (with 64 % of samples below 0.1). A few samples had negative values attributed to a high background signal. The positive values trailed between 0.1 and 4, and only two samples contained a C-circle ratio higher than 4.

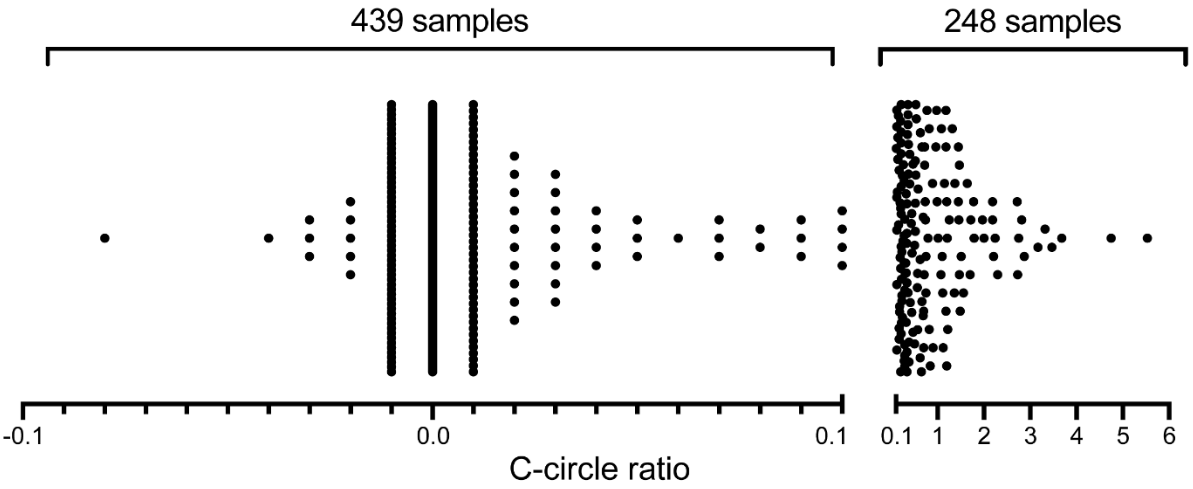


Figure 9. C-Circle distributions of analyzed samples. Scatter plot of C-circle ratio values from 687 sarcoma samples. 439 samples are shown between -0.1 and 0.1, 248 samples are binned between 0.1 and 6.

1.2 Two-gaussian distribution fit describes a heterogeneous ALT+ population

Cutoffs for the ALT+ population was determined based on normal distributions for two populations after log transformation. As previously stated, most samples had C-circle ratios around zero (Figure 10A). The C-circle values from the sarcoma samples were then transformed by adding a 0.05 pseudocount to avoid negative values that would be incompatible with the Log2 transformation. The null value, now between -6.3 and -3.5, contained the fit for the ALT- population (Figure 10B). Above -2.8 were the ALT+ cancers, with the fit encompassing all values above this threshold and reaching a maximum of 2.8. This corresponds to a variation of the normalized C-circle signal by about tenfold.

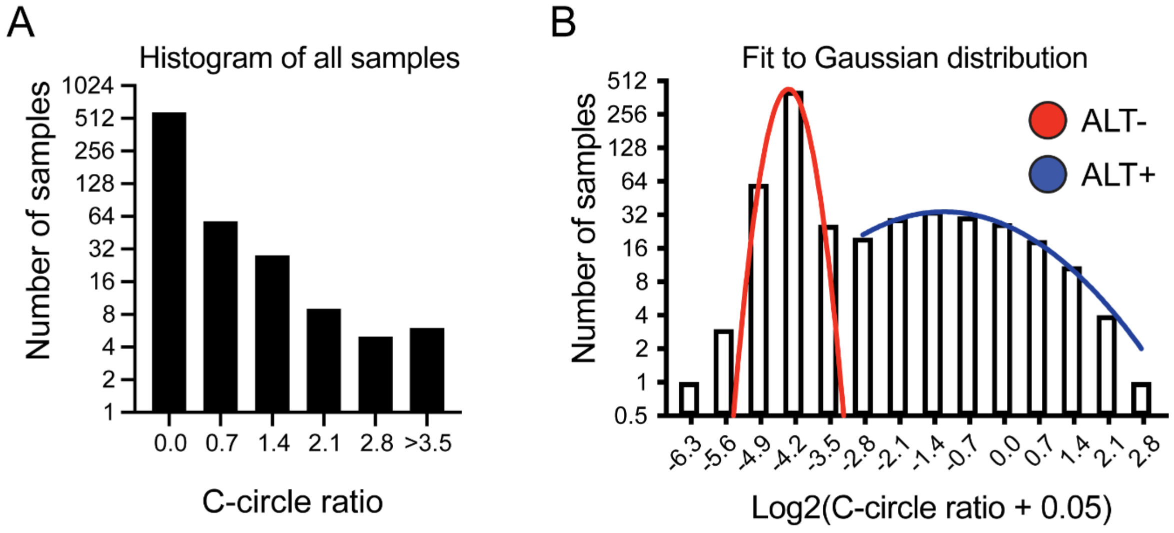


Figure 10. Data transformation and two-Gaussian fitting. A) Histogram of raw C-circle values, in bins of 0.7 with the y-axis in Log2. (B) Two-Gaussian fitting to Log2 and pseudocount transformed data, showing the ALT- population in red and ALT+ population in blue. Y-axis is in Log2.

The parameters obtained from the fit of two Gaussian distributions can be seen in Table 3. The cutoff for the ALT+ samples was determined using the 90th percentile for the ALT- and the 10th percentile for the ALT+. Accordingly, I could identify 148 samples considered ALT+ and 510 samples that were considered ALT-. The ALT+ had a mean at 0.95 normalized C-circle ratios, with a high standard deviation. This shows that the C-circle values associated with ALT+ samples were highly heterogeneous, whereas ALT- samples were almost entirely within the null value. 29 samples (4%) fit neither population cutoff and were termed "ambiguous". This population may instead represent a transition state between ALT+ and ALT- cancers; however, the low overall sample number may also be a byproduct of imposing percentile cutoffs for each population.

Table 3. Population statistics of two-gaussian distributions with number of samples

				C-circle ratio	
Population	n samples	10th percentile	90th percentile	Mean	S.D.
ALT-	510	N/A	0.069	0.00	0.01
ALT+	148	0.2	2.27	0.95	0.91
Ambiguous	29	N/A	N/A	0.12	0.03

2 Sequencing analysis

2.1 Cell line panel assembly for downstream projects

The findings mentioned in the C-circle screening section indicated that ALT-activity is highly heterogeneous in primary tumor samples. To this end, I assembled a cell line panel consisting of osteosarcoma (OS) and pediatric glioblastoma (pGBM) cell lines that were heterogeneous in ALT-activity as well as for mutations within ATRX and H3.3 (Table 4). The choice of cell lines was made using tumor entities enriched in ALT and having well-characterized ALT features. The NEM157, NEM165, NEM168 cell lines were kindly provided by Prof. Stefan Pfister (DKFZ, Heidelberg, Germany), and the NEM168 ATRX knockout clone B5 were made by Dr. Inn Chung (Division of Chromatin Networks, DKFZ, Heidelberg, Germany) using CRISPR-Cas9 (190).

Table 4. Cell line panel consisting of two tumor entities. NEM168 clone (cl.) B5 constitutes an ATRX knockout cell line with an increased C-circle ratio. C-circle ratio expressed as mean and the SEM in parentheses.

Tumor entity	Cell line	C-circle ratio	TMM status	ATRX mutations	H3.3 mutations
Pediatric glioblastoma	KNS42	0.0 (0.0)	ALT-	WT	G34V
	MGMB1	2.1 (0.6)	ALT+	Yes	G34R
	SF188	0.0 (0.0)	ALT-	WT	WT
	NEM157	0.04 (0.0)	Weak ALT+	Yes	K27M
	NEM165	0.01 (0.0)	Weak ALT+	WT	K27M
	NEM168	0.03 (0.0)	Weak ALT+	WT	K27M
	NEM168 cl. B5	0.3 (0.05)	ALT+	Yes	K27M
Osteosarcoma	HOS	0.0 (0.0)	ALT-	WT	WT
	MG.63	0.0 (0.0)	ALT-	WT	WT
	Cal72	0.7 (0.1)	ALT+	Yes	WT
	Saos-2	1.1 (0.0)	ALT+	Yes	WT
	U2OS	1.0 (0.1)	ALT+	Yes	WT

All C-circle ratios were measured with technical assistance from Lukas Frank and Caroline Knotz (Division of Chromatin Networks, DKFZ, Heidelberg, Germany). Samples were binned in ALT- (0 C-circles), weakly ALT+ (0.01-0.1 C-circles), or ALT+ (>0.1 C-circles). Mutations within ATRX and H3.3 were either obtained by genomic analysis by Dr. Katarina Deeg (Division of Chromatin Networks, DKFZ, Heidelberg, Germany) for the pediatric glioblastoma cell lines (190), or were previously published for the osteosarcoma cell lines (191). ILSE-identification numbers from the DKFZ core facility for each sequencing readout and cell line replicate are found in Supplemental Table 1.

Long and small RNA-seq, together with ATAC-seq, were performed in duplicates for each cell line. An overview of the sequencing and respective downstream analysis is found in (Figure 11). The long RNA-seq fraction with both protein-coding and non-coding genes is referred to here as the transcriptome. and thus active enhancers and promoters, offered a unique perspective in observing Some larger transposable elements (TE) can also be found in this data (192). The small RNA-seq provided information on miRNA as well as TE and their silencers piRNA. The ATAC-seq, which includes information on open chromatin regions processes not necessarily found in transcriptomic data. Altogether, the latter sequencing readouts have not been performed previously from an ALT-specific viewpoint.

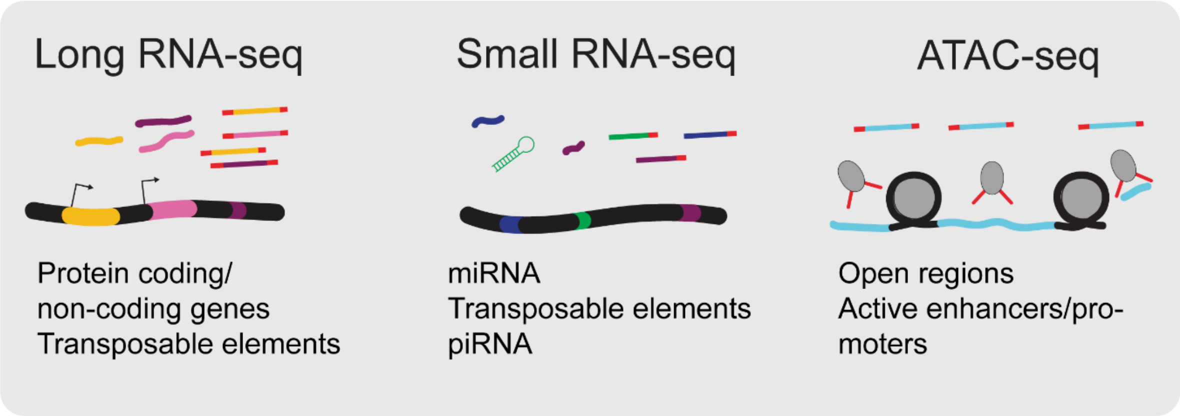


Figure 11. Overview of the sequencing workflow and downstream analyses.

The downstream computational analysis was performed by observing differential expression with DESeq2. The first step consisted of evaluating which genes were differentially expressed and what transcription factor (TF) motifs were found in up or downregulated genes. The differentially expressed genes were annotated by gene ontology and overlapping these genes with the ALT gene database Telnet (185). Moreover, pathway analysis was performed on transcriptomic data. miRNAs were also analyzed for differential expression. However, due to limited curated miRNA annotations, the miRNA analyses were mainly evaluated using literature search and databases with gene-target predictions. Transposable elements (TE) and piRNA were evaluated for differential features as well. Due to limited annotation databases, these elements were only tested for whether ALT+ samples contained significant differences. ATAC-seq provided information on the epigenomic state of the ALT+ chromatin where differentially accessible peaks were assigned to genomic elements, such as promoters or intergenic regions, and analyzed for TF motifs. Altogether these individual readouts provided information on which elements were the most accessible from sequencing data and could then be passed on to multi-omics integration. The first step consisted of selecting ALT gene candidates by overlapping differential ATAC peaks with promoters and enhancers and then connecting these to differential ALT genes. This further homed in on the potential biomarkers that could be analyzed in the future. Lastly, I investigated whether an ALT signature could be extracted from all the individual omics using an unsupervised multi-omics factor analysis (MOFA) (174). MOFA allowed discerning more discrete data patterns and answered whether an ALT-signature is readily found throughout the sequencing data or if it could be confounded by other factors (e.g., sequencing run or cell line gender).

2.2 Eigencorrelation analysis shows that tumor entity is a confounding factor in omics data

The first step of the omics data analysis was to evaluate the presence of ALT and any confounding cell line metadata. The analysis was done using principal component analysis (PCA) and performing correlations between each component (PC) and cell line metadata. The metadata provided was cell line gender, sequencing run, tumor entity, telomere maintenance mechanism (TMM), and the cell line itself. It was found that the TMM was primarily present in PC1, which would be the component that can explain most of the variance in the data (Figure 12). However, the TMM often co-occurred with the tumor entity. This is best seen for transcriptome, miRNA, piRNA and ATAC. It was therefore decided to normalize for tumor entity. This was performed for the differential analysis of individual readouts by including the tumor entity within the design factor for DEseq2 (see Materials and Methods section 6.6). For multi-omics factor analysis (MOFA), this was done by using the batch correction functionality from Limma (see Materials and Methods section 8.3).

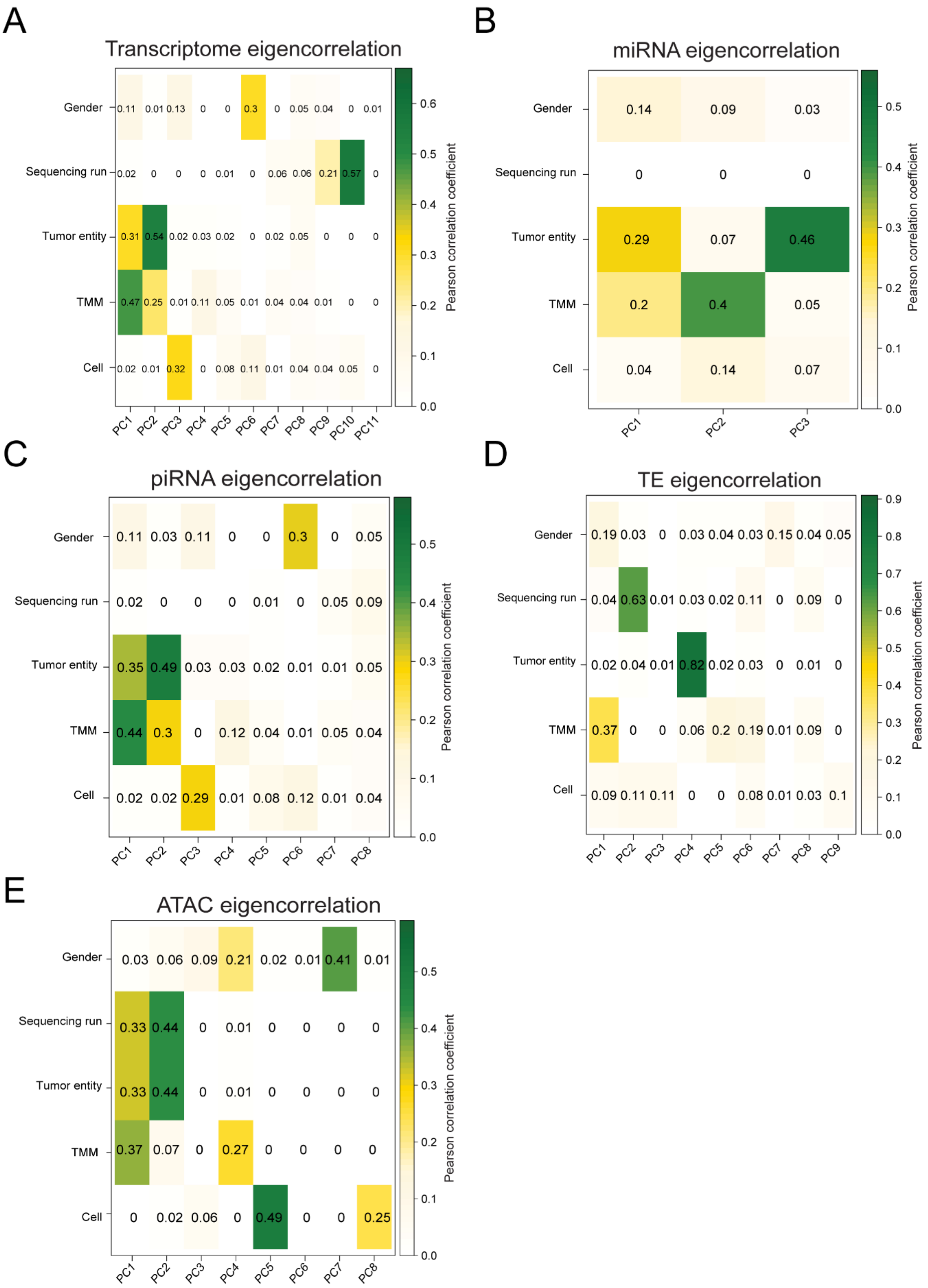


Figure 12. Eigencorrelation between cell line metadata and principal components.

3 Individual sequencing read-outs

3.1 Transcriptome quality control

The quality control for the long RNA-seq was determined by investigating how many reads were aligned to the hg38 genome build. This metric consisted of at least 100 million reads, which was achieved for each sample (Figure 13).

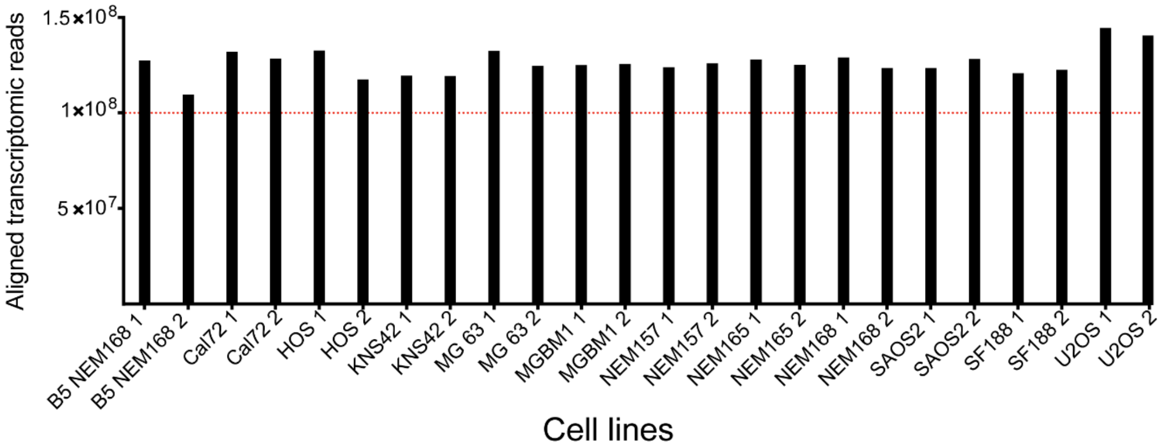


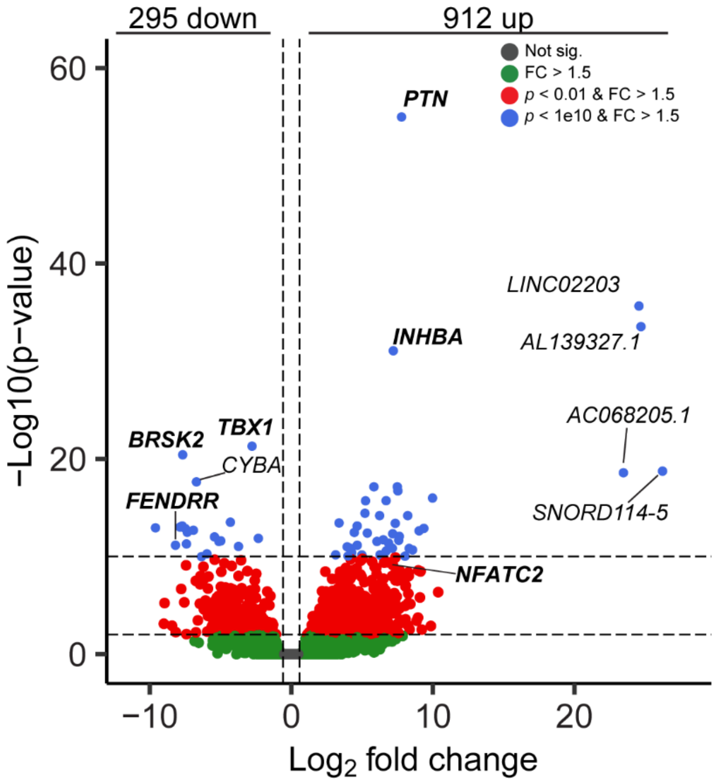
Figure 13. Quality control of aligned transcriptomic reads.

3.2 Gene expression changes are robustly identified in ALT+ samples

Out of the 64,915 ENSEMBL annotated genes, 37,848 genes passed the low-read cutoff. The resulting DESeq2 analysis revealed 295 downregulated and 912 upregulated genes between ALT- and ALT+ cell lines (Figure 14). Additionally, 78 genes (of which 54 were upregulated and 24 downregulated) had an adjusted p-value below $1 \cdot 10^{-10}$ (see Supplemental Table 2 for complete gene list). The differential genes were then selected to find consistent expression changes in the ALT+/ALT- cell lines. The top consistently downregulated genes were the *BRSK2* kinase, the *TBX1* transcription factor, and the long noncoding RNA *FENDRR*. The top consistently upregulated genes were *PTN*, a secreted growth factor, and *INHBA*, a membrane protein subunit related to activin and inhibin signaling. Another consistently upregulated gene was the transcription factor *NFATC2*, which was close to the high *p*-value cutoff. The long noncoding RNA *LINC02203*, the novel transcripts *AL139327.1* and

AC068205.1, and the small nucleolar RNA (snoRNA) *SNORD114-5* had very high *p*-values and fold changes. However, these were highly upregulated in only a subset of ALT+ samples.

Figure 14. Volcano plot of ALT+ differentially expressed genes. Red indicates peaks that passed both fold change (1.5) and adjusted p-value cutoffs (0.01). Green shows peaks that only passed fold change cutoff, and grey shows peaks below both cutoffs. Blue dots show the differential genes with an adjusted p-value above $1{\cdot}10^{10}$. *295 downregulated and 912 upregulated genes could be identified above statistical and fold-change cutoffs. Genes in bold denote expression levels that were consistently up/down in most ALT+/ALT- samples.*



3.3 Transcription factor motif analysis of differentially expressed genes indicates that inflammation upregulates genes

The differentially expressed genes were then used for transcription factor (TF) motif enrichment. The enrichment was calculated 10 kb upstream of the transcription start site. For the downregulated genes, the top motif was associated with the TFs Egr-1 and Egr-2 (Figure 15). The Egr TFs work downstream of NF-kB signaling pathways to promote apoptosis (193). The second top motif was the Max/c-Myc motif. This is of particular interest, as c-Myc is believed to upregulate TERT in ALT- cancers (194). As this TF motif is associated with both c-Myc and Max, there is a chance that this c-Myc/Max dimer is particularly crucial in inhibiting ALT. Another downregulated TF motif found in the downregulated ALT differentially expressed genes was Hic1, which is a tumor suppressor that can maintain genetic stability (195). The upregulated genes contained immune-related TFs, such as STAT1 (alpha and beta isoforms) 3, 6, and NFAT1/2. STAT 1,3 and 6 are activated by interferons, which are inflammatory cytokines (196), and NFATs are TFs that regulate T-cell response (197). NFAT2 is the protein name for *NFATC2* gene, which was found consistently upregulated in ALT transcriptome (Figure 14).

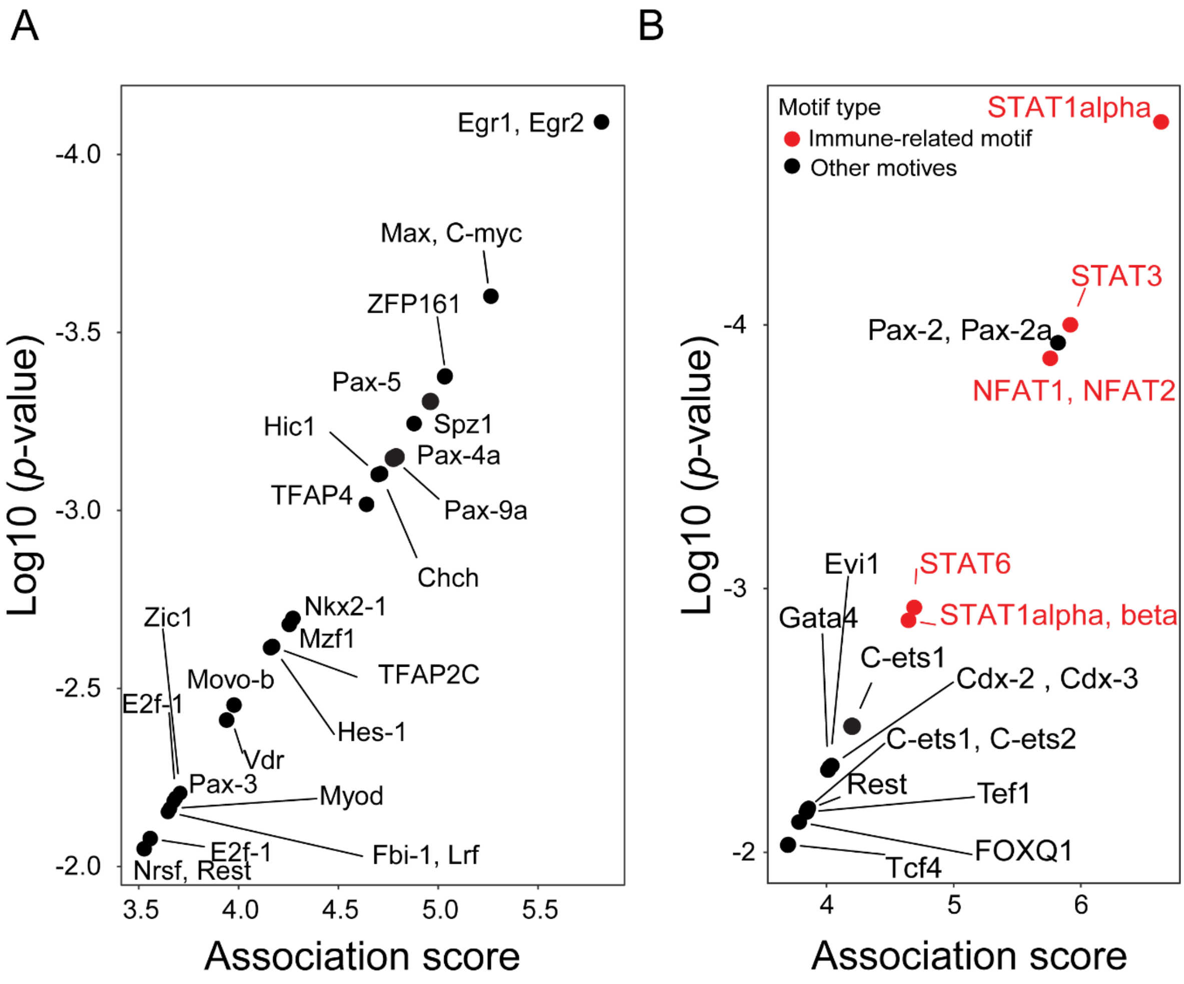


Figure 15. Transcription factor motif analysis in promoter regions of differentially expressed ALT genes. Motifs with a p-value lower than 0.01 were omitted. Y-axis is the Log10 p-value plotted, and x-axis shows the association score which is calculated from the -Log10 hypergeometric p-value. (A) Transcription factor motifs in downregulated genes. (B) Transcription factors in upregulated genes. Red denotes transcription factors involved in immune processes.

3.4 GO-term analysis of differential genes reveals changes within developmental programs

The differentially expressed genes were divided into upregulated and downregulated genes and then used for GO-term analysis to identify biological processes. Most of the terms were connected to developmental processes (Table 5). Of note was the negative enrichment in proteins related to nucleic acid metabolic processes, indicating that this process is noticeable absent in upregulated ALT+ differential genes. Positive enrichment for biological/cell adhesion and cell differentiation were also found, showing a relationship between upregulated genes and cancer cell invasion and motility. Lastly, the GO-term "generation of neurons" was found as well.

Table 5. Top 10 GO-terms associated with upregulated ALT+ differentially expressed genes.

GO-term ID	GO-term	Enrichment	FDR
GO:0022610	Biological adhesion	+	4.42E-10
GO:0007155	Cell adhesion	+	1.35E-09
GO:0032501	Multicellular organismal process	+	1.06E-08
GO:0090304	Nucleic acid metabolic process	-	3.03E-08
GO:0048731	System development	+	2.26E-07
GO:0048869	Cellular developmental process	+	3.95E-07
GO:0048856	Anatomical structure development	+	3.95E-07
GO:0032502	Developmental process	+	4.35E-07
GO:0030154	Cell differentiation	+	4.42E-07
GO:0048699	Generation of neurons	+	4.49E-07

The downregulated GO-terms show an overlap with the upregulated GO-terms, consisting of anatomical structure development, system development, cellular development process, and multicellular organismal process (Table 6). This implies that rather than relying on specific genetic programs, the ALT+ differential genes are involved in fine-tuning developmental programs beneficial to the tumors.

Table 6. Top 10 GO-terms associated with downregulated ALT+ differentially expressed genes.

GO-term ID	GO-term	Enrichment	FDR
GO:0009653	Anatomical structure morphogenesis	+	7.66E-04
GO:0007417	Central nervous system development	+	9.34E-04
GO:0032502	Developmental process	+	9.40E-04
GO:0048513	Animal organ development	+	9.98E-04
GO:0032501	Multicellular organismal process	+	1.07E-03
GO:0048856	Anatomical structure development	+	1.09E-03
GO:0007275	Multicellular organism development	+	1.26E-03
GO:0030154	Cell differentiation	+	1.35E-03
GO:0048731	System development	+	2.10E-03
GO:0048869	Cellular developmental process	+	2.22E-03

3.5 Annotation with the Telnet database points to diverse functions in ALT

The differential genes were then annotated using the Telnet database to delve further into which ALT+ genes play an essential biological role. Telnet is a manually curated database that consists of ALT-related genes and proteins in both yeast and human cells (185). The overlap yielded 50 genes and constituted a plethora of functions relating to cell biology. Among them were the two top downregulated genes, *TBX1* and *BRSK2*, and the upregulated gene *NFATC2* (Figure 16).

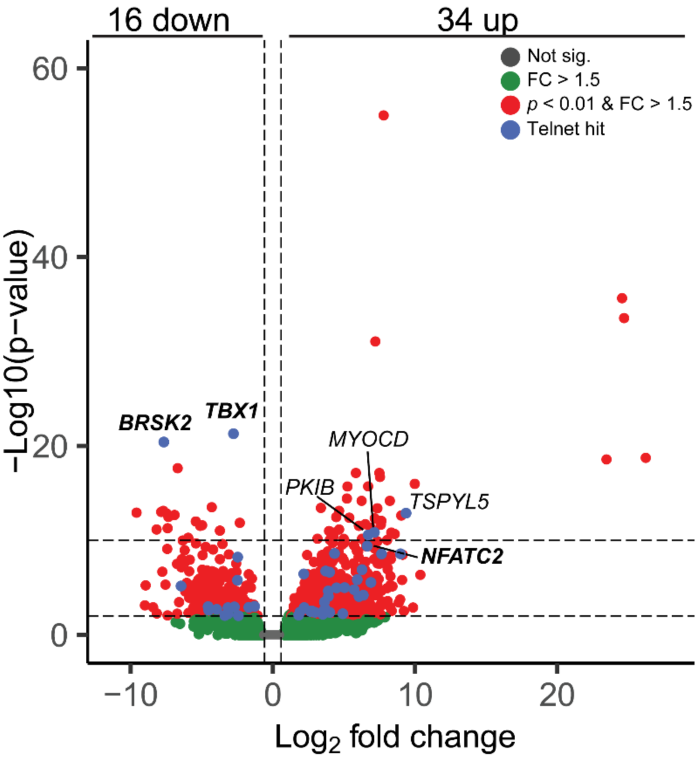


Figure 16. Volcano plot of differentially expressed genes in ALT+ samples. Red indicates genes that passed both fold change (1.5) and adjusted p-value cutoffs (0.01). Green shows genes that only passed fold change cutoff, and grey shows genes below both cutoffs. Blue shows genes annotated in the Telnet database. Genes in bold denote expression levels that were consistently up/down in most ALT+/ALT- samples.

The full list of Telnet annotated genes is shown in Table 7. The Telnet annotations revealed that *BRSK2* was identified as a positive telomerase regulator in an RNAi kinase screen (198), whereas *TBX1* was found in pulldown experiments using shelterin components (199). Another three genes in the high p-value group were annotated within the Telnet database, consisting of *MYOCD*, *PKIB*, and *TSPYL5*. Of note, however, was that these genes were highly expressed in some ALT+ cell lines, but not the majority. The transcription factor *NFATC2,* while not belonging to the high p-value group, had consistent upregulation in almost all ALT+ cell lines, found as a TF in upregulated genes (see Results section 3.3), and was also annotated in the Telnet database.

NFATC2 was annotated as a telomerase-enhancing hit as its motif is found in *TERT* promoters (200). *HIC1* encodes the Hic1 TF which was found in downregulated ALT genes, and its Telnet annotation shows that it found deregulated in ALT+ cell lines (119).

Table 7 Differential ALT+ genes annotated in the Telnet database. Bold denotes genes that were consistently downregulated/upregulated in most cell lines.

Gene name	Function	Gene name	Function
TDRD6	Cell differentiation	***BRSK2***	Protein kinase
TDRKH	Cell differentiation	*CAMK1G*	Protein kinase
TSPYL5	Cell growth	*MAPK15*	Protein kinase
ANOS1	Cell membrane	*PKIB*	Protein kinase
ARL4C	Cell membrane	*PPP1R17*	Protein kinase
ELOVL7	Cell membrane	*HIC1*	RNA transcription and processing
IFITM1	Cell membrane	*MOV10L1*	RNA transcription and processing
ABCC9	Cell membrane	*MYOCD*	RNA transcription and processing / TF
MAGEA4	Cell membrane	***NFATC2***	RNA transcription and processing / TF
PAGE2	Cell membrane	*PARM1*	RNA transcription and processing
PAGE5	Cell membrane	*ESR1*	RNA transcription and processing / TF
LRRC63	Cell membrane	*ARPP21*	RNA-binding
HDAC9	Chromatin organization	*ARHGAP15*	Signaling
HIST1H4F	Chromatin organization	*NLRP2*	Signaling
KMT2A	Chromatin organization	*PLCD1*	Signaling
SETDB1	Chromatin organization	*FOXP2*	TF
CDK15	Cyclin kinase	*FOXR1*	TF
CDKL2	Cyclin kinase	*GATA2*	TF
ANKLE1	DNA replication and repair	*GATA5*	TF
LRATD1	DNA replication and repair	*KLF12*	TF
AKR1B10	Metabolism	*KLF8*	TF
ASS1	Metabolism	*SOX5*	TF
EPHX2	Metabolism	***TBX1***	TF
SULT4A1	Metabolism	*TFEC*	TF
TNIP3	Metabolism		

3.6 Reactome analysis of ALT+ samples finds neuronal and developmental pathways

Another approach for functionally determining gene expression changes in ALT+ samples was made by pathway analysis. Count tables from transcriptome alignment were used, and replicates were merged by averaging gene expression. These were then submitted for pathway analysis using the down-weighting of overlapping genes (PADOG) method (201). Using PADOG resulted in increased sensitivity of the gene set enrichment analysis. The Reactome pathway enrichment was then calculated, and the top 10 pathways were extracted in terms of Q-value (Table 8).

Table 8. Top 10 Reactome pathways in ALT+ cell lines.

Reactome pathway	Directionality	Q-value
Signaling by activin	Up	2.0E-05
Antagonism of activin by follistatin	Up	2.0E-05
Other semaphorin interactions	Up	2.0E-05
NGF-independent TRKA activation	Down	2.0E-05
Neuropilin interactions with VEGF and VEGFR	Up	2.0E-05
VEGF ligand-receptor interactions	Down	2.0E-05
VEGF binds to VEGFR leading to receptor dimerization	Down	2.0E-05
RUNX1 regulates estrogen receptor mediated transcription	Up	2.0E-05
RUNX1 regulates transcription of genes involved in differentiation of myeloid cells	Up	2.0E-05
RUNX1 regulates transcription of genes involved in WNT signaling	Up	2.0E-05

These showed a relationship between activin and follistatin signaling, in which the top upregulated gene *INHBA* (see Results section 3.2) is a key factor of inhibiting activin signalling. Neuronal signaling pathways relating to semaphorins and NGF-independent TRKA activation were found, where the former was upregulated, and the latter was downregulated. Angiogenesis also seemed to be disrupted, where neuropilin interactions with VEGF/R were upregulated, yet VEGF ligand-receptor interactions and VEGF/R receptor dimerization were downregulated. This illustrates that ALT shows a specific preference for the VEGF ligands and the neuropilin receptor, bypassing the more canonical VEGF receptors. While neuropilin is physiologically relevant in neuronal outgrowth, cancer cells can also use it to regulate angiogenesis (202). RUNX1, a TF whose motif was found in upregulated ATAC peaks (see Results section 3.12), had upregulated pathways related to estrogen receptor transcription, myeloid differentiation, and WNT signaling. Of note is the identical Q-value for these top 10 pathways, where the low sample size of the 12 averaged cell lines most likely influenced the binning used for the Q-values.

3.7 Quality control parameters for miRNA-seq pipeline

To investigate the influence of miRNA in ALT+ samples, a new miRNA analysis pipeline was implemented. The minimum number of aligned reads for each sample was 3 million, which was achieved for all replicates (Figure 17A). The following quality metric concerned how many miRNA hairpins were detected, with a maximum of 1,551 known hairpins . The cutoff was set at 1,000 hairpins (Figure 17B), where each replicate attained this goal. In this case, the reason for choosing hairpins is due to the biology and computational analysis that is generally performed for miRNA. miRNAs are generated as miRNA hairpins, which are then cleaved into mature miRNAs that are biologically functional (203).

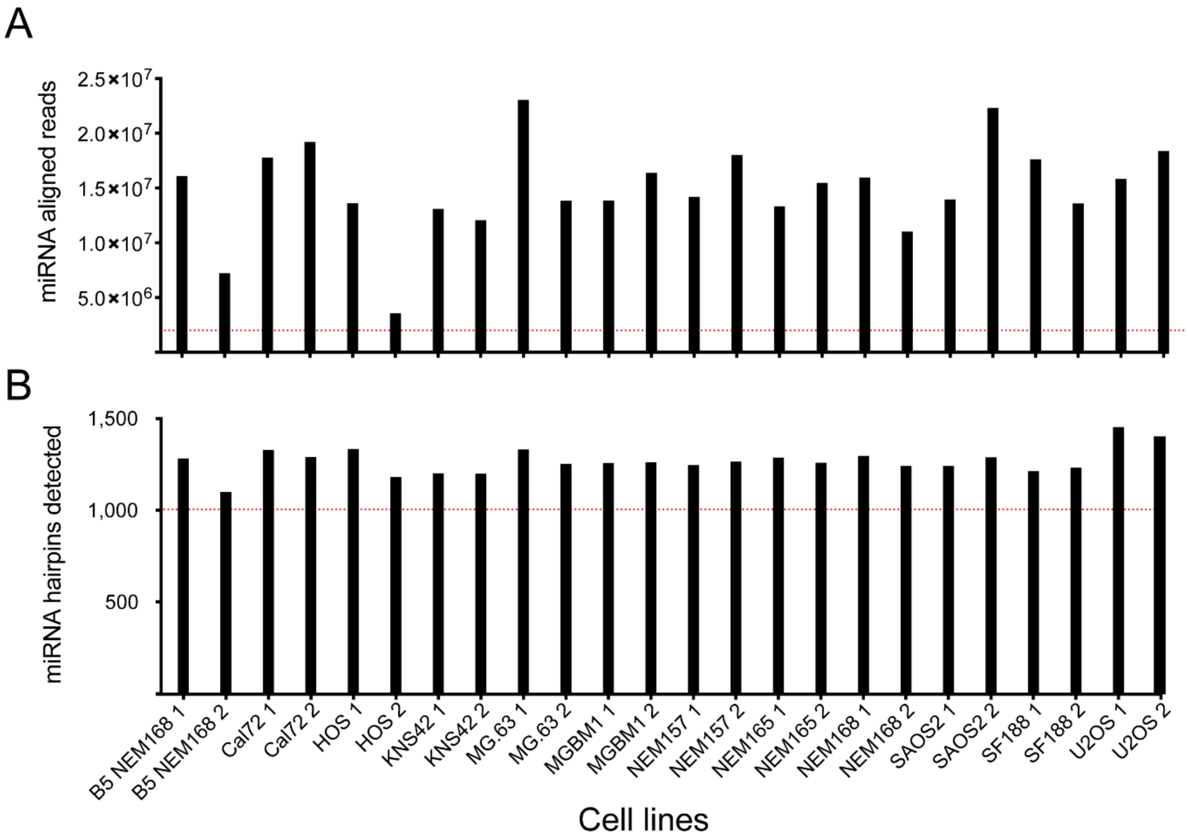


Figure 17. Quality control parameter for miRNA pipeline. (A) Aligned reads in hundred million. (B) miRNA hairpins detected.

3.8 Identification of differentially expressed miRNAs show telomere maintenance-related hits

The mature miRNA reads were used to construct count tables and analyzed with DEseq2. When passing low-read filtering, 822 out of 2,425 mature miRNAs were used for differential expression. A total of 52 miRNA were identified as differentially expressed (Figure 18), with 8 downregulated and 44 upregulated miRNAs. More upregulated features than downregulated features were found. The corresponding miRNA IDs can be found in Table 9. Of the 52 differentially expressed miRNA, 17 had GO-term “negative regulation of telomeres” based on which mRNA they can regulate.

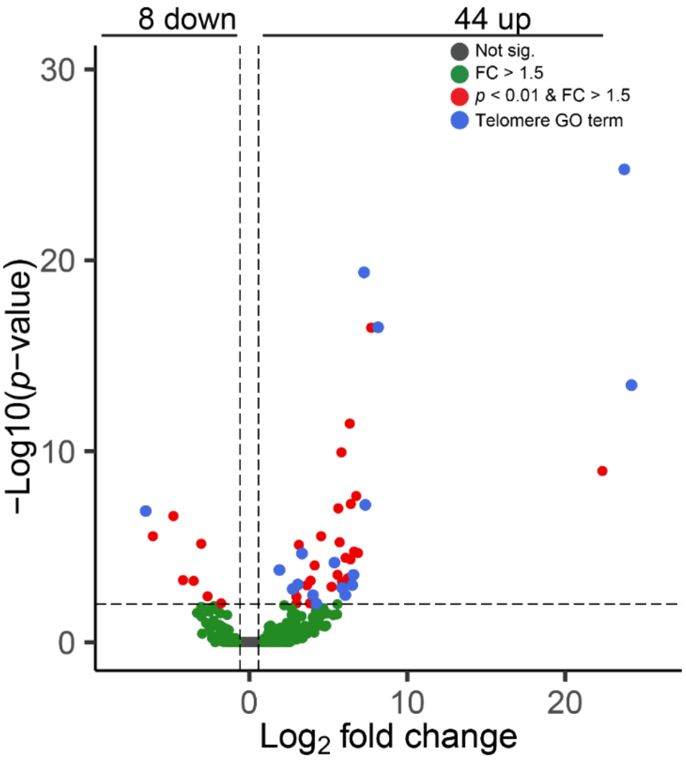


Figure 18. Volcano plot of ALT+ differentially expressed miRNA. Red indicates peaks that passed both fold change (1.5) and adjusted p-value cutoffs (0.01). Green shows peaks that only passed fold change cutoff, and grey shows peaks below both cutoffs. Blue shows the 17 miRNAs with the gene ontology (GO) term “negative regulation of telomere elongation”. 2,425 mature miRNAs passed low read cutoffs, where 8 downregulated and 44 upregulated miRNAs could be identified above statistical and fold-change cutoffs.

Table 9. All differentially expressed miRNA, divided into downregulated (downreg.) and upregulated (upreg.). miRNAs associated with telomere-related gene ontology terms are denoted in bold.

Downreg. miRNA	Upreg. miRNA	Upreg. miRNA cont.
hsa-miR-199b-5p	hsa-miR-34a-5p	**hsa-miR-490-3p**
hsa-miR-214-3p	hsa-miR-105-5p	**hsa-miR-512-3p**
hsa-miR-214-5p	hsa-miR-105-3p	hsa-miR-516a-5p
hsa-miR-3120-3p	hsa-miR-122b-5p	**hsa-miR-519a-3p**
hsa-miR-3120-5p	**hsa-miR-122-5p**	hsa-miR-519a-2-5p
hsa-miR-4677-5p	**hsa-miR-129-5p**	hsa-miR-519d-5p
hsa-miR-6716-3p	**hsa-miR-129-1-3p**	hsa-miR-520b-5p
hsa-miR-10399-5p	**hsa-miR-129-2-3p**	hsa-miR-520f-5p
	hsa-miR-138-5p	hsa-miR-522-3p
	hsa-miR-138-1-3p	hsa-miR-708-3p
	hsa-miR-139-5p	**hsa-miR-708-5p**
	hsa-miR-139-3p	hsa-miR-767-3p
	hsa-miR-142-3p	**hsa-miR-767-5p**
	hsa-miR-146a-3p	hsa-miR-876-3p
	hsa-miR-204-3p	hsa-miR-1284
	hsa-miR-204-5p	hsa-miR-3117-3p
	hsa-miR-218-5p	hsa-miR-3167
	hsa-miR-323a-3p	hsa-miR-3909
	hsa-miR-323b-3p	hsa-miR-4652-5p
	hsa-miR-372-3p	hsa-miR-4662a-5p
	hsa-miR-375-3p	hsa-miR-5579-3p
	hsa-miR-377-5p	hsa-miR-5579-5p
	hsa-miR-485-5p	

3.9 Investigating transposable elements and piRNA

Next, transposable elements (TEs) and their piRNA silencers were investigated. Count tables were generated from aligned reads with an in-house RNA-seq pipeline, allowing multimapping reads (see Materials and Methods section 6.14). Reads mapping to exonic counts were excluded. While 1,180 TE families were detected, none were differentially expressed (Figure 19A). This might be due to TEs being comprised of thousands of different transcripts that were grouped into families (such as LINE-1) (204). Thus, individual transcript changes were not represented within their family. Another option was that the TEs were already being silenced to the extent that they were not differentially expressed. To support the latter statement, two downregulated, and 12 upregulated piRNAs were be found (Figure 19B). While future analysis with piRNA target prediction would be beneficial in answering what specific TE transcripts are targeted, this result still indicates that piRNAs are differentially expressed and of interest in ALT.

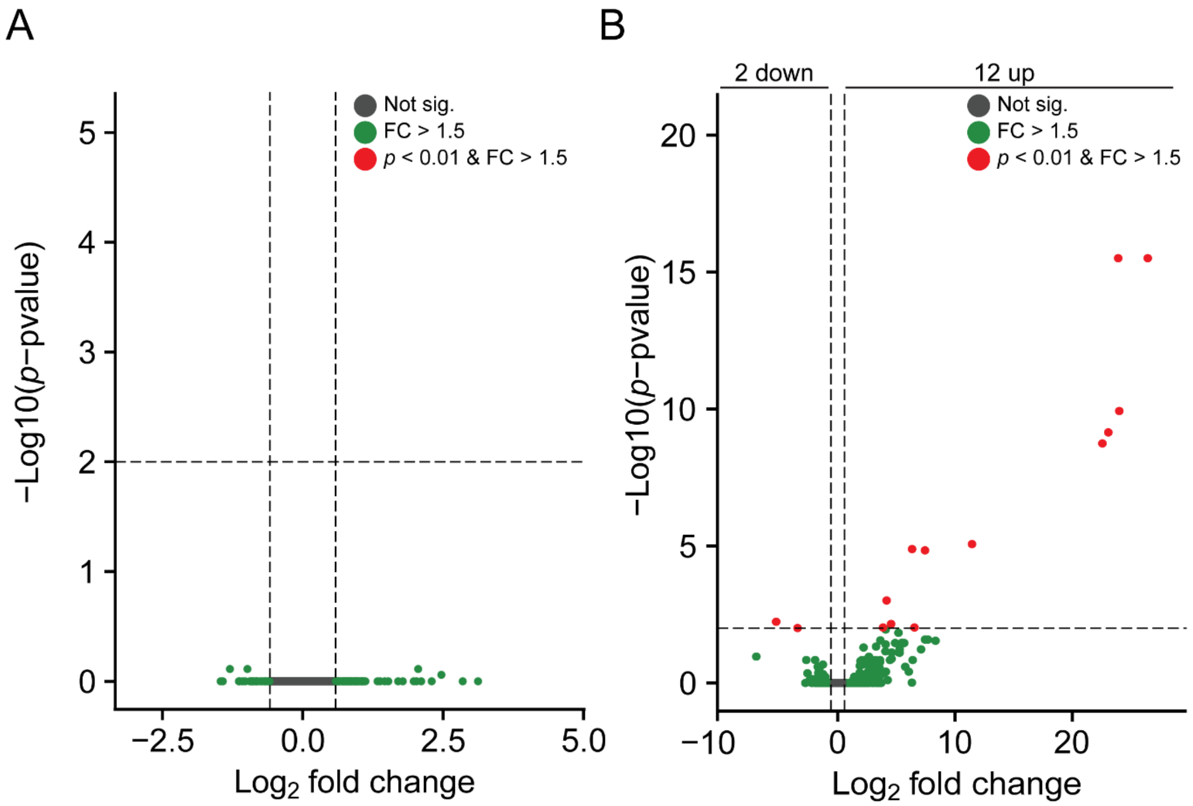


Figure 19. Differential expression of transposable elements (A) and piRNA (B). 1,180 transposable element families and 1,926 piRNAs were detected. Red indicates features that passed both fold change (1.5) and adjusted p-value cutoffs (0.01). Green shows features that only passed fold change cutoff, and grey shows features below both cutoffs.

3.10 Quality control parameters for ATAC-seq pipeline

The samples were sequenced at a read depth of approximately 50 million reads using paired-end sequencing. A quality cutoff at a minimum of 20 million aligned reads was used, and Figure 20A shows that this was achieved. The second parameter was how many peaks could be identified, where a minimum of 50,000 peaks was to be expected. Again, this was achieved as seen in Figure 20B. The last important metric analyzed was the Fractions of Reads in Peaks (FRiP), which indicated the fraction of ATAC reads were found in an identified peak. The FRiP cutoff was put at 0.1, which all samples passed (Figure 20C).

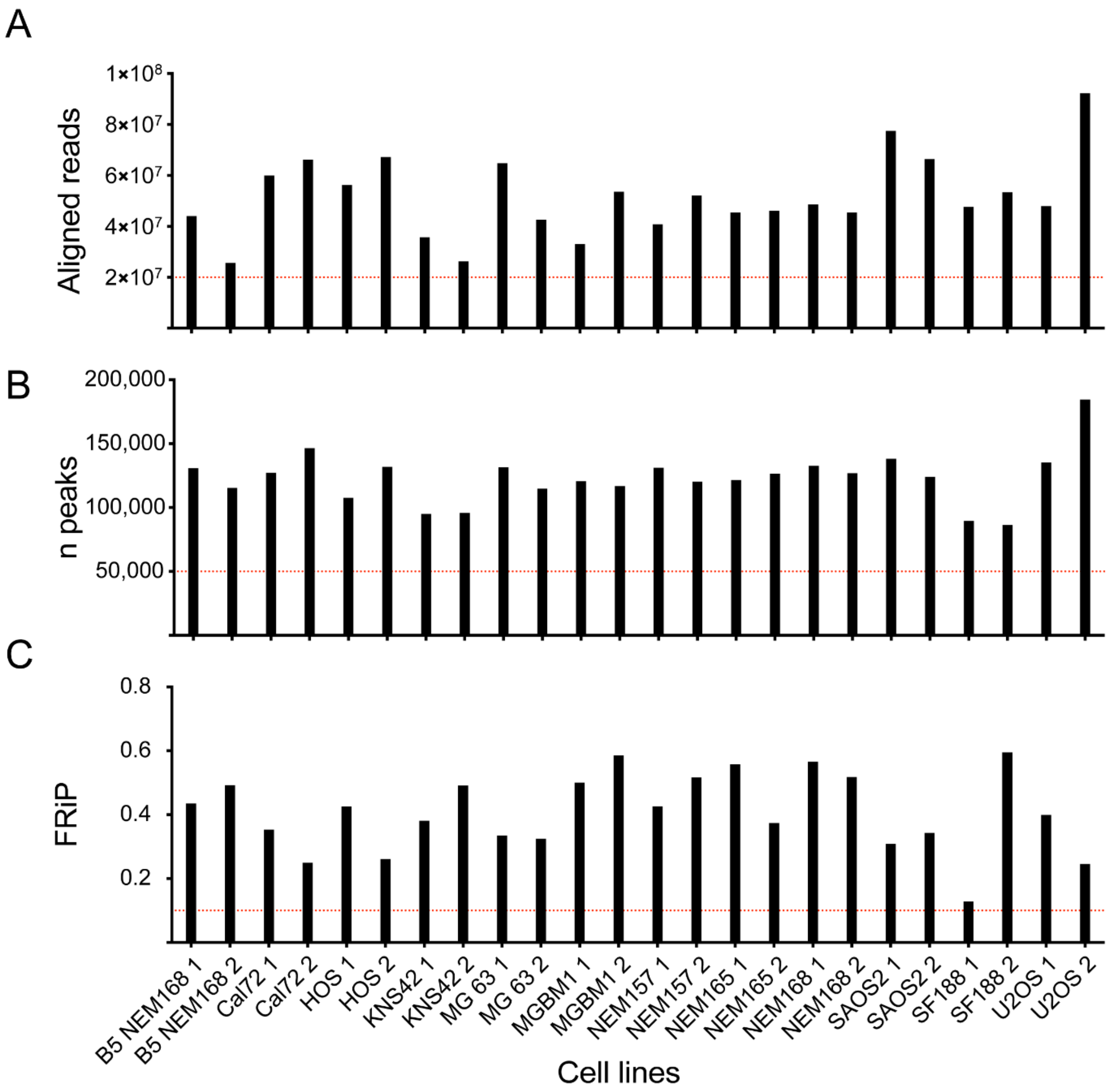


Figure 20. Quality control parameters for ATAC-seq pipeline. (A) Aligned reads in 10 million. (B) The number of identified peaks through peak calling. (C) FRiP score for each replicate.

3.11 Differential expression analysis identifies thousands of differential peaks mainly in the noncoding genome

A count table was constructed against the consensus peak set created from the ATAC-seq pipeline, and the Y-chromosome was removed to avoid gender bias. The resulting count table contained 471,337 peaks. The corresponding volcano plot is shown in Figure 21A, with 2,015 downregulated and 3,418 upregulated peaks above the cutoff criteria. More upregulated peaks were found than downregulated ones, indicating a general opening of chromatin in ALT+ cells. The differential peaks were then annotated to their genomic location (Figure 21B). Most of the peaks were found in non-coding regions, such as intronic and distal intergenic (83%). Only 9 % were found in promoters, and 6 % were found in exons. This suggests that most of the ALT+ changes in chromatin accessibility occurred within non-coding regions.

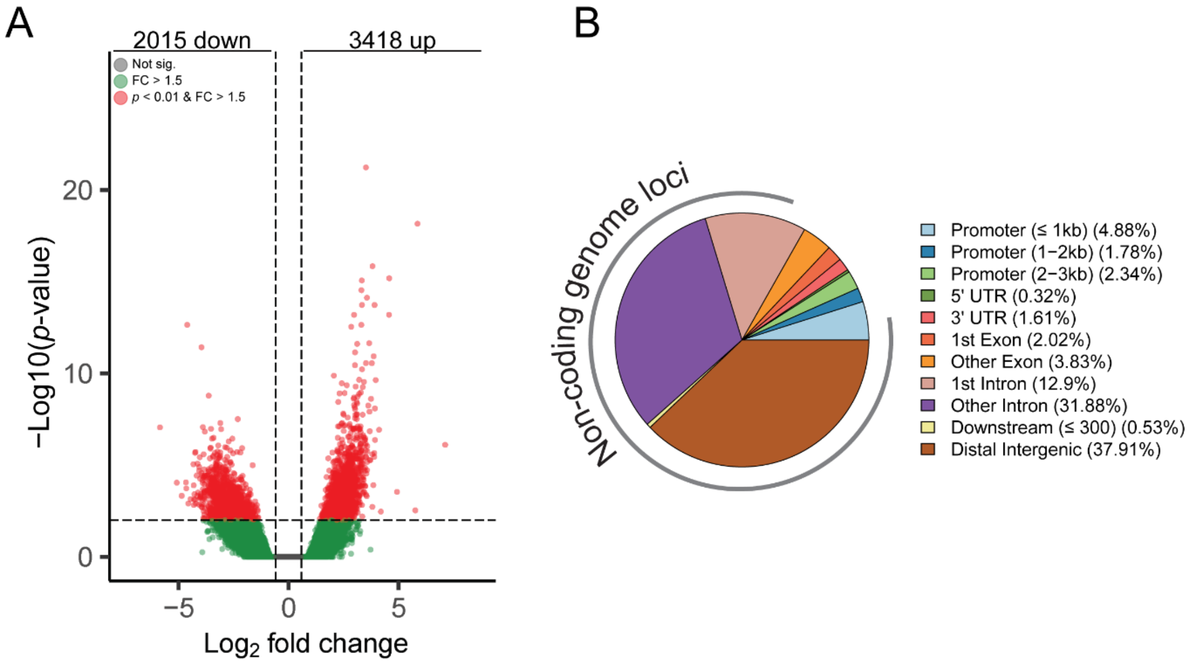


Figure 21. Differential accessibility in ATAC peaks and annotation. (A) Volcano plot of differentially accessible peaks in ALT+ samples. Red indicates peaks that passed both fold change (1.5) and adjusted p-value cutoffs (0.01). Green shows peaks that only passed fold change cutoff, and grey shows peaks below both cutoffs. 471,337 peaks passed low-read cutoffs, 2,015 downregulated, and 3,418 upregulated peaks were identified above statistical and fold-change cutoffs. (B) Annotation of differentially expressed peaks.

3.12 ATAC-peaks are enriched in developmental and cell-survival transcription factor families

The differentially accessible peaks were then analyzed for TF motifs. A cutoff *p*-value of $1\cdot10^{60}$ was used, and additionally, the enrichment was calculated by dividing the % of motifs in the background *versus* the % of motifs in target sequences. For the downregulated peaks, almost all the top TF motifs in terms of *p*-value belonged to TFs in the SOX family (Figure 22A). The upregulated peaks were identified to contain members in the RUNX and AP-1 family motifs exclusively (Figure 22B).

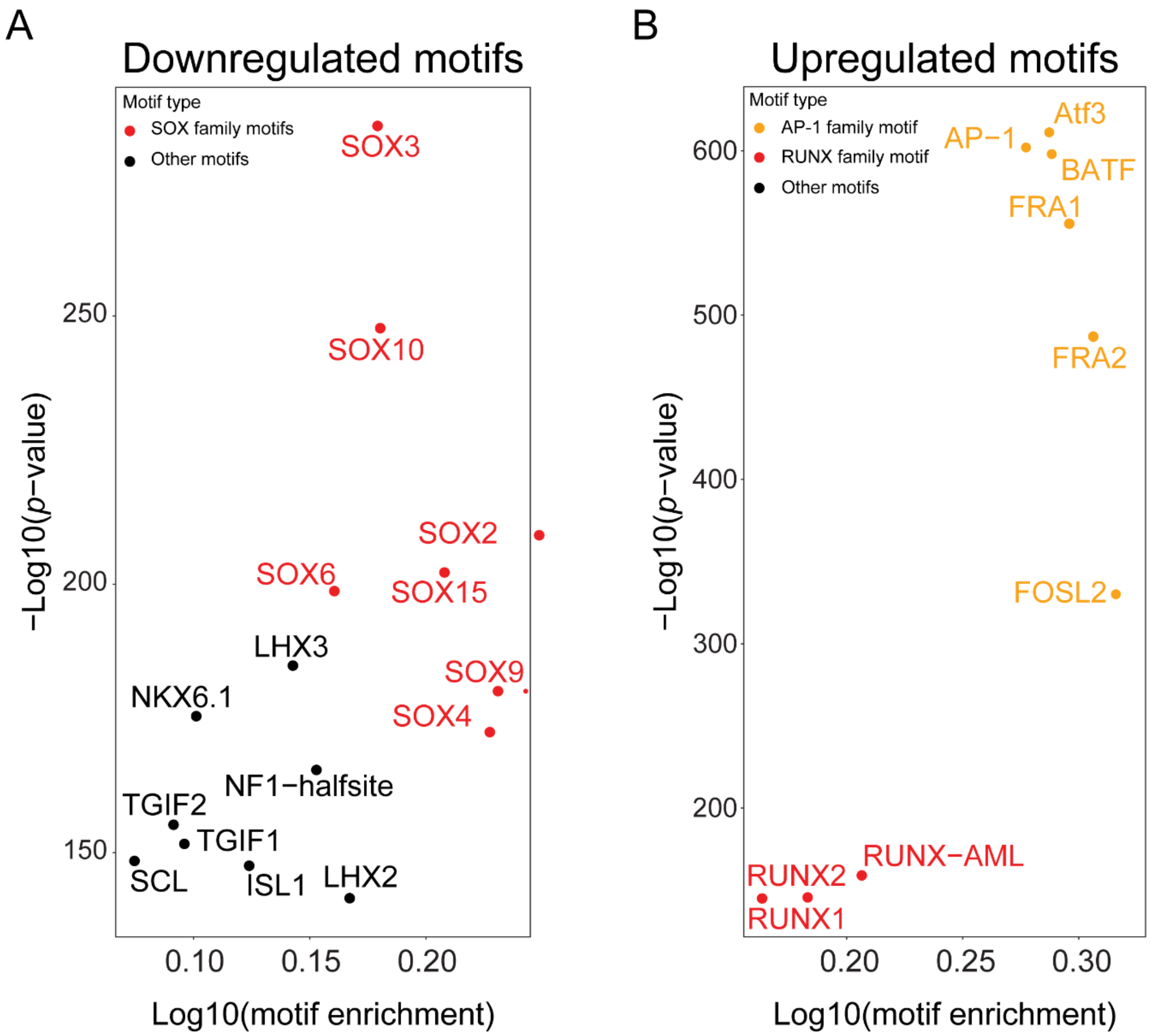


Figure 22. Transcription factor motifs in ALT+ differentially accessible peaks. Shown with the y-axis as the -Log10 p-value and the x-axis as the -Log10 motif enrichment. (A) Motifs in downregulated peaks, with SOX family motifs in red. (B) Motifs in upregulated peaks, with RUNX family motifs in red and AP-1 family motifs in yellow.

Both SOX and RUNX are involved in developmental processes (205, 206), yet the distinct downregulation of SOX and upregulation of RUNX indicate that the chromatin state of ALT+ cell lines specifically favors RUNX but not SOX. Additionally, RUNX-AML motifs were also found. This motif is based on RUNX1 mutations in hematological cancers and correlates with poor prognosis in acute myeloid leukemia (207). *SOX5*, while being found as downregulated in transcriptomic data (Table 7), was not present in the TF motif analysis. This is most likely be due to gene expression not being a complete determinant of TF activity and TFs are extensively regulated via post-translational modifications (208). The AP-1 family has a broad function relating to cell survival and is commonly deregulated in cancer (209). It has also been shown that the AP-1 transcription factors are connected to inflammation (210). This may be particularly relevant as TF motif analysis of differentially expressed genes showed that upregulated genes contained many immune-related TF motifs (Figure 15). Therefore, it could point towards gene expression programs in ALT being influenced by immune signaling pathways.

4 Integrative omics analysis

4.1 ATAC and RNA integration reveal two potential ALT biomarkers

To investigate whether differentially accessible peaks and expressed genes showed signs of coregulation, differential ATAC peaks were divided into promoters or enhancers for genes. The ENCODE database defined the promoter region, ordinarily close to the transcription start site of a gene. These promoter sites were then expanded to include 1 kb upstream or downstream of the promoter sites. The enhancers were defined according to ENCODE annotated enhancers, and each activate enhancer was connected to the nearest gene. These active promoters and enhancers were further linked to differentially expressed genes from section 3.2 and the Telnet database by overlapping gene names. For the upregulated genes and peaks, 10 hits could be found (Figure 23A). 142 differentially expressed genes contained an upregulated promoter/enhancer but without being annotated to a gene within the Telnet database. Two hits were observed for the downregulated genes and peaks (Figure 23B). 13 genes with active promoters/enhancers outside the Telnet were also found. The ATAC and RNA-seq integration further reinforced the notion that *NFATC2* and *BRSK2* genes are potential ALT biomarkers. For *NFATC2,* most ALT+ cell lines showed an increase in normalized counts, except for the U2OS cell line (Figure 24A). Nevertheless, its expression was virtually absent in ALT- cell lines (Figure 24C). Three differentially accessible peaks were found 15 kb upstream of the gene, and these showed a relationship with the gene expression. Notably, the increased accessibility was not seen for the U2OS cell line, which did not contain high levels of *NFATC2* transcript. *BRSK2* was the top downregulated gene for ALT+ samples and contained one differentially accessible peak was identified 13 kb upstream of the gene. Its expression was almost absent in all ALT+ samples (Figure 24C).

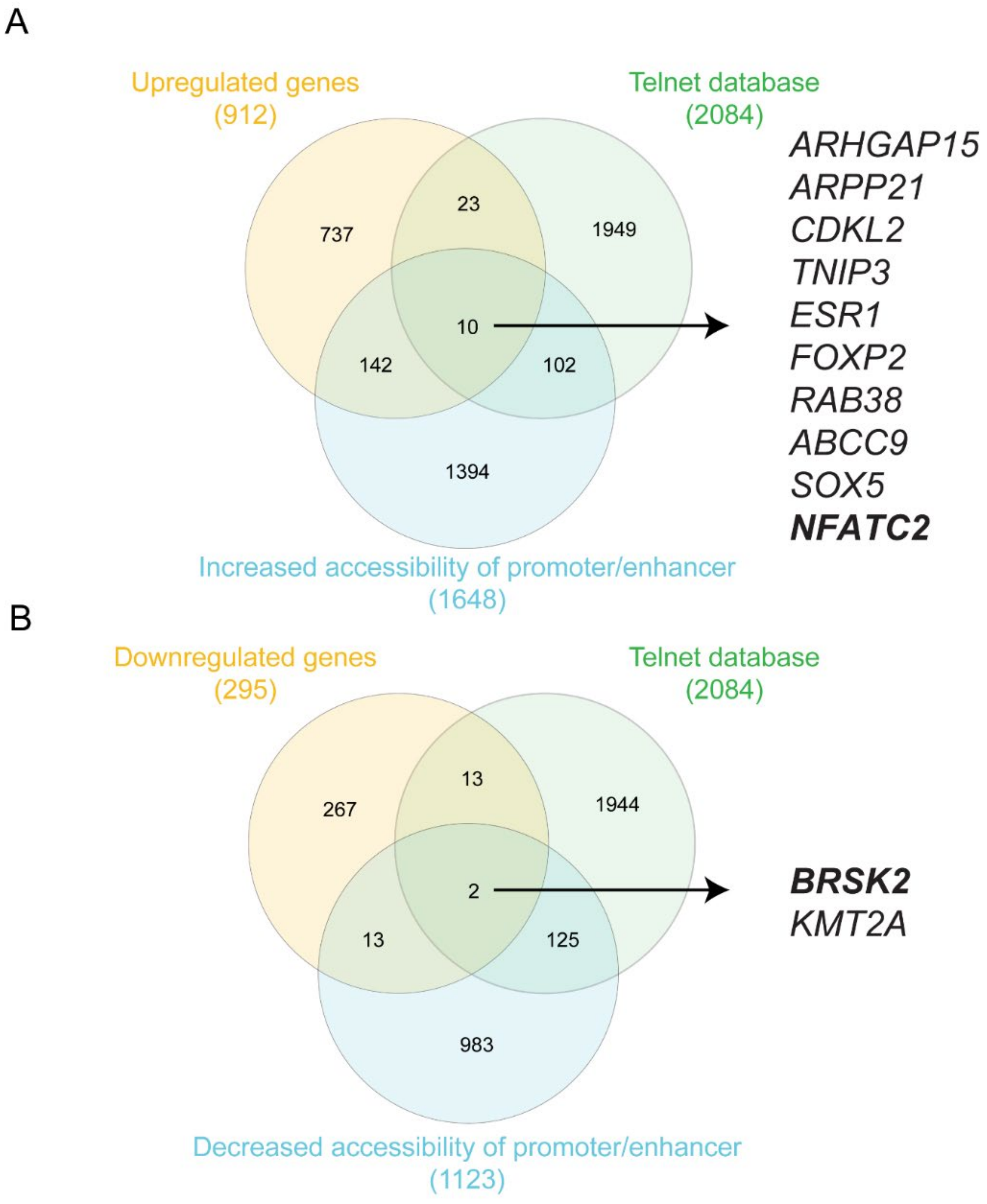


Figure 23. Results from overlaps between gene expression, promoter/enhancer ATAC signal, and Telnet database. Bold shows consistently up/downregulated genes in most cell lines (A) Upregulated gene and promoter/enhancers overlap with Telnet. (B) Downregulated gene and promoter/enhancers overlap with Telnet.

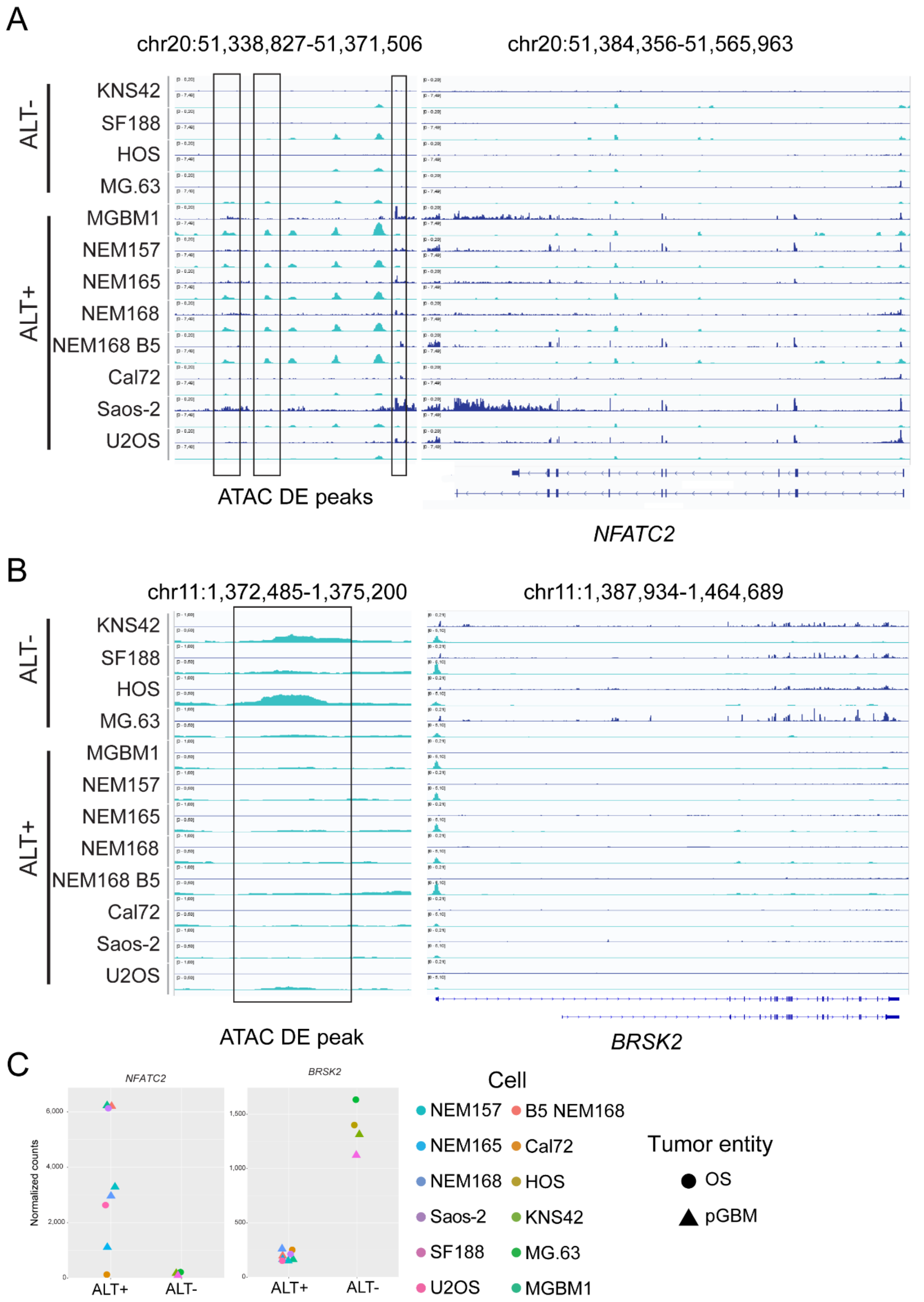


Figure 24. Examples of differentially expressed genes in ALT+ samples with differential ATAC peaks. (A) NFATC2, (B) BRSK2. Gene tracks on the right show ATAC (cyan) and RNA (blue). Differentially expressed ATAC peaks are shown on the leftmost panel with the protein-coding gene on the right. (C) Normalized counts of the NFATC2 and BRSK2 gene, each cell line is color-coded according to legend, with OS as circles and pGBM cells as triangles.

4.2 Multi-omics integration reveals a robust ALT signature

To find an ALT+ signature spanning all the different sequencing approaches, a multi-omics integration was performed. Count tables from expressed genes, ATAC, miRNA, TE, and piRNA, were transformed using variance stabilizing transformation and normalized for cancer entity. These count tables were sorted on features with the highest standard deviation in the samples and used for multi-omics factor analysis (MOFA). The analysis parameters were optimized by choosing how many variable features should be used and limiting how many features should be contained within the signature (Figure 25A). Irrespective of the number of features, TMM was the principal factor of variance in the data. The number of features analyzed was chosen based on the best distinction between ALT+ and ALT- groups, which was 1,000 features. Subsequently, the number of analyzed factors was limited to 6 factors based on the elbow method (Figure 25B). The elbow method consisted of observing the point of diminishing returns to explaining the variance when adding more factors.

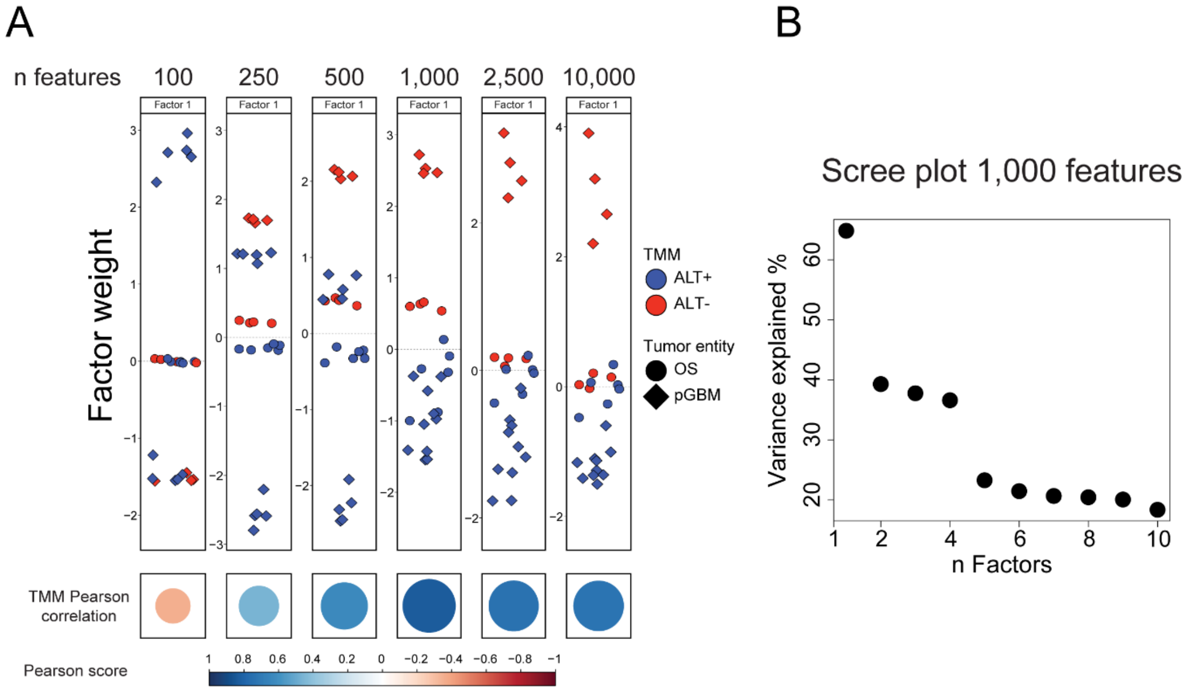


Figure 25. Optimization of multifactor omics analysis. (A) Beeswarm plots of a varying number of features. Y-axis denotes the factor weight of each cell line. Colored in red for ALT+ and blue as ALT- cell lines. Circles are OS and diamonds as pGBM. Pearson correlation of factor to TMM is shown for each beeswarm plot, with red showing negative correlation and blue positive correlation expressed as a Pearson score. The size of the dots also relates to the Pearson score. (B) Scree plot of % variance explained for a cumulative number of factors. The total number of explained variance is 100 %.

The resulting analysis was then further characterized by testing the correlation between cell line metadata and factors. Factor 1 was heavily linked to TMM, with smaller correlations between cell-cell heterogeneity or cell gender (Figure 26A). As expected, when normalizing the count tables, no correlation could be seen between the factors and tumor entity. The other factors show a mix of correlations with no apparent relation, except for factor 6, which was strongly linked to sequencing run. The Pearson correlation coefficient was extracted between each factor to observe whether factor 1 was influenced by other factors (Figure 26B).

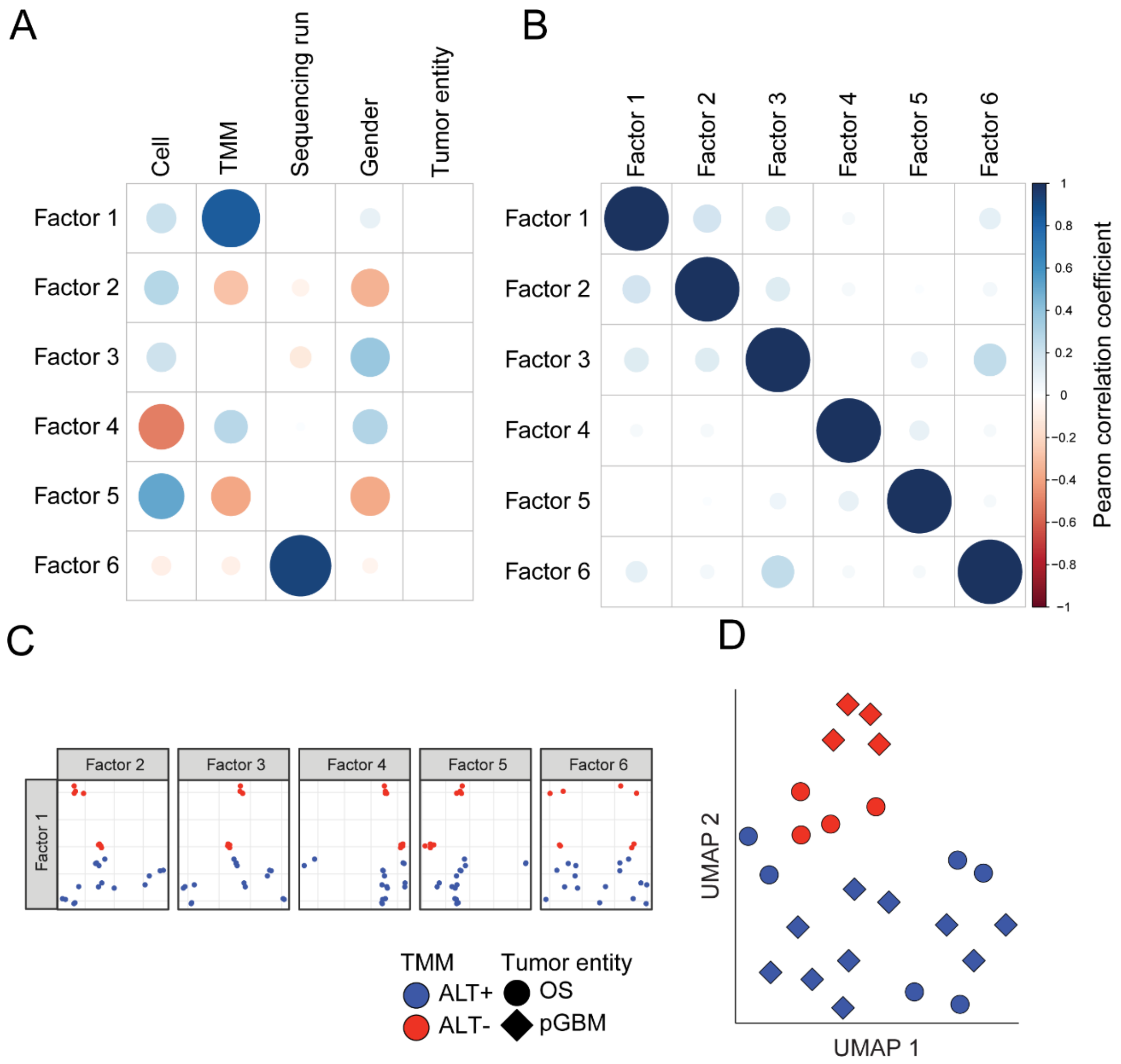


Figure 26. Testing robustness of telomere maintenance mechanism (TMM) factor. (A) Pearson correlation between factors and cell line data. Red shows negative association, and blue shows positive association. (B) Factor-factor Pearson correlation. (C) Factor-factor scatter plots with ALT+ cells in red and ALT- cells in cyan. (D) UMAP constructed of all factors, ALT+ shown in blue and ALT- in red. Circles are OS and diamonds as pGBM.

Some correlation can be seen between factors 2,3 and 6; however, the correlations were minor (> 0.4). The distinction between ALT+ and ALT- cell lines was always present, even when comparing factor 1 to other factors (Figure 26C). The only noticeable difference in sample clustering occurred in factor 6, which further separated the groups into sequencing run. Lastly, uniform manifold approximation and projection (UMAP) visualization of all factors showed the distinction between ALT- and ALT+ was ever-present in the data (Figure 26D), further indicating that the principal source of variance in the data stemmed from TMM and with little interference from other factors. As such, it was concluded that MOFA robustly extracted an ALT signature throughout the omics. Notably, the ALT signature was the principal source of variation and was not confounded by other factors.

4.3 The ALT gene expression signature is linked to oxidative stress and cell motility

I then performed a functional analysis of the ALT signature. The functional analysis consisted in testing what omics contribute most to the signature. In this case, transcriptome described most of the variance (27%), followed by miRNA (19%) and ATAC (14%) (Figure 27A). piRNA and TE only explained 9 % and 7 % of the variance, respectively. This shows that ALT+ samples could be best identified through transcriptome, miRNA, and chromatin accessibility. As gene expression was the most significant determinant of ALT+ in omics data, GO-terms were extracted for the negatively and positively weighed genes for factor 1. It is to be emphasized that the genes that were negatively weighted for factor 1 were instead positively weighted for ALT+ samples, as ALT+ samples themselves were negatively weighted for factor 1 (Figure 25A). GO-terms related to cellular response to alcohol, cell motility, and neurogenesis were seen for the positively weighted genes. The last two terms are also represented within the GO-term analysis from section 3.4. The identical *p*-values most likely originated from the same genes annotated to all five pathways (Supplemental Figure 1B).

The response to reactive oxygen species was highly enriched for the negatively weighted genes, with other GO-terms denoting cochlea development, retinoic acid response, and negative neuron death regulation. Reactive oxygen species may relate to intrinsic mitochondrial dysfunction and drive pro-survival pathways through cell signaling (211). This was further indicated by the "response to retinoic acid" GO-term, where ALT cell lines may modulate fatty acid metabolism to either counter or potentiate oxidative stress (212). Additionally, genes annotated in negative regulation of neuron death may be related to countering reactive oxygen species induced apoptosis, and cellular protein complex assembly may be related to ER stress (213, 214).

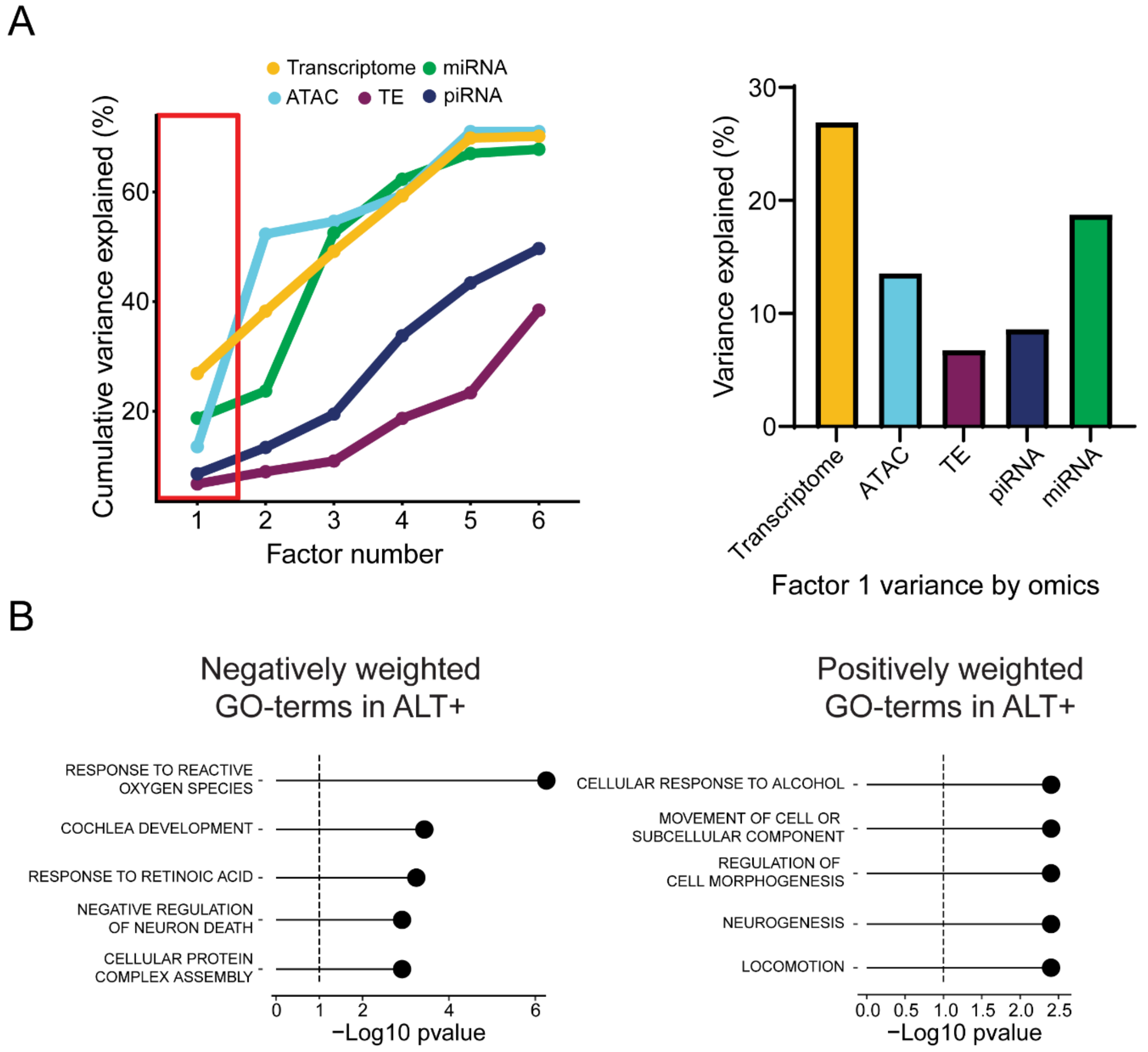


Figure 27. MOFA factor 1 characterization. (A) Variance explained between each omics and factor. The leftmost plot shows the cumulative variance between all the factors and all the omics, and the rightmost plot shows the variance explained by each omics in factor 1. (B) GO-term analysis from negatively or positively weighted genes in ALT.

5 Perturbing epigenetic marks in ALT with drugs

5.1 ALT+ viability is affected by inhibition of DNA methylation and H3K27 trimethylation

To investigate whether ALT cells showed susceptibilities for chemical inhibitors, I assembled 15 chemical compounds that were of potential ALT interest and were considered specific for their targets at the time (Table 10). These compounds targeted DNA repair, epigenetic processes, innate immunity sensing, and telomerase. Mitotic kinases were also screened to determine non-lethal concentrations used for metaphase spreads (see Results section 5.6). DNA repair was previously established as a potential ALT+ therapeutic (215), while epigenetic inhibitors have not been tested previously from an ALT perspective. Innate immunity sensing was chosen based on the finding that C-circles induced DNA-sensing pathways via the cGAS/STING signaling pathway (216). To this end, an inhibitor for cGAS and two STING agonists were used. The latter of which encompassed the natural cGAS second messenger cGAMP and the small molecule G10.

Table 10. List of compounds used to find ALT-specific susceptibility or resistance in terms of cell viability. Bold denotes positive hits.

Compound name	Target
5-azacytidine	**DNA methylation**
AZD7648	DNA-PK
CC1 pan-kinase inhibitor	Broad kinase inhibitor
BIBR 1532	Telomerase
cGAMP	cGAS-STING second messenger
CP2	KDM4
CX-5461	rRNA synthesis
G10	STING (agonist)
GSK343	**EZH2 mediated H3K27me3**
IKK-16	IKK
KU-55933	ATM
LDN-192960	Haspin
LY2603618	Chk1
RU.521	cGAS
ZM 447439	AURKA/B

A reduced screening format was utilized to find compounds that showed differences in ALT+ cell viability. This was done using two cell lines, KNS42 as an ALT- control and MGBM1 as an ALT+ control. The compounds that showed noticeable cell viability differences were then subjected to cell viability tests with the entire cell line panel consisting of eight pediatric glioblastoma cell lines (Table 11). These cell lines were the same as those used in the sequencing analysis (Table 4), with the addition of the NEM168 ATRX knockout clone F2. The pGBM cell lines constituted a variety of mutations in H3.3 and ATRX and different ALT activities. From the 14 compounds screened, two were investigated further. These were 5-azacytidine, which targeted DNA methylation, and GSK343 which targeted PRC2 mediated H3K27me3 deposition. It is noted that at the time of the experiment, the cGAS inhibitor RU.521 was novel and believed to be specific to human cGAS (217). However, subsequent studies have shown that RU.521 is mainly active on murine cGAS and much less so for human cGAS (218). Still, due to the negative results for the STING agonists G10 and cGAMP, it was concluded that ALT+ cell lines were not highly sensitive to disruption of DNA sensing.

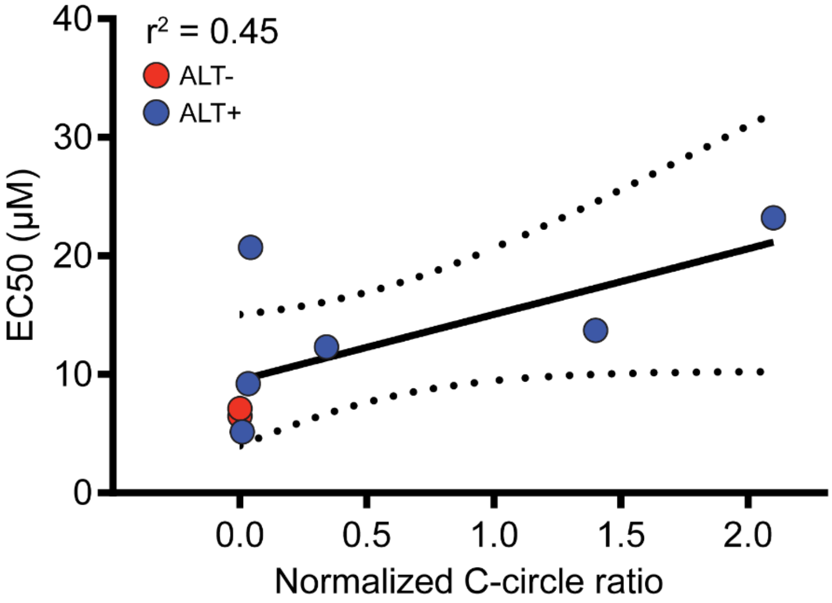
Table 11. Cell line panel used in inhibitor experiment. NEM168 clones B5 and F2 constitute ATRX knockouts where ALT activity was increased. C-circle ratios shown as mean and SEM in parentheses. µM EC_{50} is shown as means with standard deviation in parentheses. EC_{50} results stem from three biological replicates consisting of three technical replicates.

Tumor entity	Cell line	C-circle ratio	TMM	ATRX mutation	H3.3 mutation	Aberrant H3.3s31p	µM EC_{50}	
							5-Azacytidine	GSK343
Pediatric glioblastoma	KNS42	0 (0.0)	ALT-	WT	G34V	No	7.1 (0.85)	7.9 (1.6)
	SF188	0 (0.0)	ALT-	WT	WT	No	6.5 (0.78)	9.6 (1.6)
	MGBM1	2.1 (0.6)	ALT+	Yes	G34R	Yes	23.2 (0.70)	5.9 (1.9)
	NEM157	0.04 (0.0)	Weak ALT+	Yes	K27M	Yes	20.7 (0.75)	6.4 (1.3)
	NEM165	0.01 (0.0)	Weak ALT+	WT	K27M	No	5.1 (0.77)	10.3 (1.8)
	NEM168	0.03 (0.05)	Weak ALT+	WT	K27M	No	9.2 (0.78)	10.3 (1.8)
	NEM168 cl. B5	0.3 (0.0)	ALT+	Yes	K27M	No	12.3 (0.77)	9.6 (1.6)
	NEM168 cl. F2	1.4 (0.6)	ALT+	Yes	K27M	No	12.7 (0.85)	5.1 (1.6)

5.2 5-azacytidine treatment shows a weak correlation between ALT activity and cell viability

The first hit was the DNA methylation inhibitor 5-azacytidine. A trend was observed regarding cell lines with higher C-circle values being more resistant to 5-azacytidine (Figure 28). While most cell lines with low (weakly ALT) or no discernable C-circles (ALT-) had low EC_{50} values when compared to samples binned to higher C-circle values (ALT). One exception was the NEM157 cell line, which although being binned as weakly ALT, had EC_{50} values comparative to the MGBM1 cell line (which had the highest C-circle ratios of all cell lines) (see values in Table 11).

Figure 28. EC_{50} for cell survival in two ALT- cell lines (red) and 6 ALT+ cell lines (blue). The X-axis shows the normalized C-circle ratios for each cell line, and the Y-axis the effective concentration at which 50 % of cell signal is lost. The sample size consists of three technical replicates repeated three times. r^2 denotes the correlation coefficient.



5.3 GSK343 is heat-labile which affects downstream applications

The heat lability for GSK343 was determined due to initial findings showing highly irregular cell viability results. DMSO or 16 µM GSK343 were pre-incubated by mixing with cell medium and incubated at 37 °C at 5 % CO_2. These were in turn added to the MGBM1 cell line (Figure 29). As is evident, at 16 µM GSK343 the cell viability decrease was absent when the compound was pre-incubated for three days. For the cell viability assay, the GSK343 compound was therefore reapplied every three days. While it is noted that even at one-day pre-incubation, roughly 25 % of cells survive, a repeated medium change caused cell detachment, especially in high-dose conditions where the cells were already stressed.

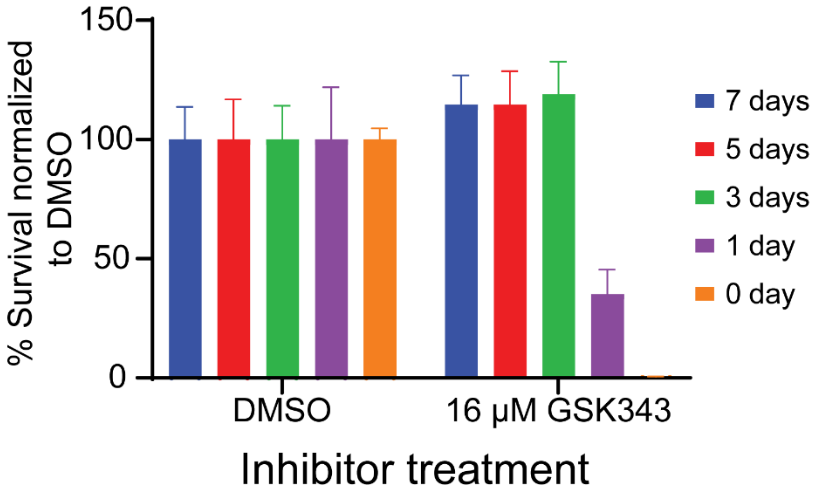


Figure 29. Heat lability test for GSK343 in the MGBM1 cell line. Y-axis shows % survival normalized to DMSO. Average of three technical replicates.

5.4 GSK343 shows a weak correlation between ALT activity and cell viability

GSK343 was then applied to the full cell line panel after the optimization described above. A general trend was seen in cell lines with higher ALT-activity being more susceptible to EZH2 inhibition (Figure 30A). In juxtaposition to a previously published study (219), this susceptibility was not connected to H3K27M mutants. As seen in Table 11, the NEM cell lines binned to weakly ALT+, which contained H3K27M substitutions, did not have markedly different viability than ALT- samples. Instead, the MGBM1 and NEM168 ATRX -/- clone F2, which had the highest C-circle ratios, were more susceptible. This indicated that EZH2 inhibition targets cell viability in cells with higher ALT activity. The next question was in which cell death modality the compound reduced the cell viability signal.

By performing a cell death flow cytometry assay, I tested whether apoptosis (Annexin V-FITC positive populations) and necrosis (propidium iodide positive populations) were different between DMSO and cell lines exposed to 6 µM GSK343 (Figure 30B). KNS42 and SF188 were chosen as ALT- controls, MGBM1 as a high ALT activity control, and NEM157 as a weakly ALT+ sample. While being in an inhibitor concentration well within the range of the EC_{50} values, the cell lines did not show a significant increase in any kind of cell death. The KNS42 cell line showed a slight increase in apoptotic/necrotic cells; however, this value was around 3 %. For the other cell lines, while having error bars spanning -5% to +5% cell death increase, the conclusion was still drawn that GSK343 inhibits cell proliferation.

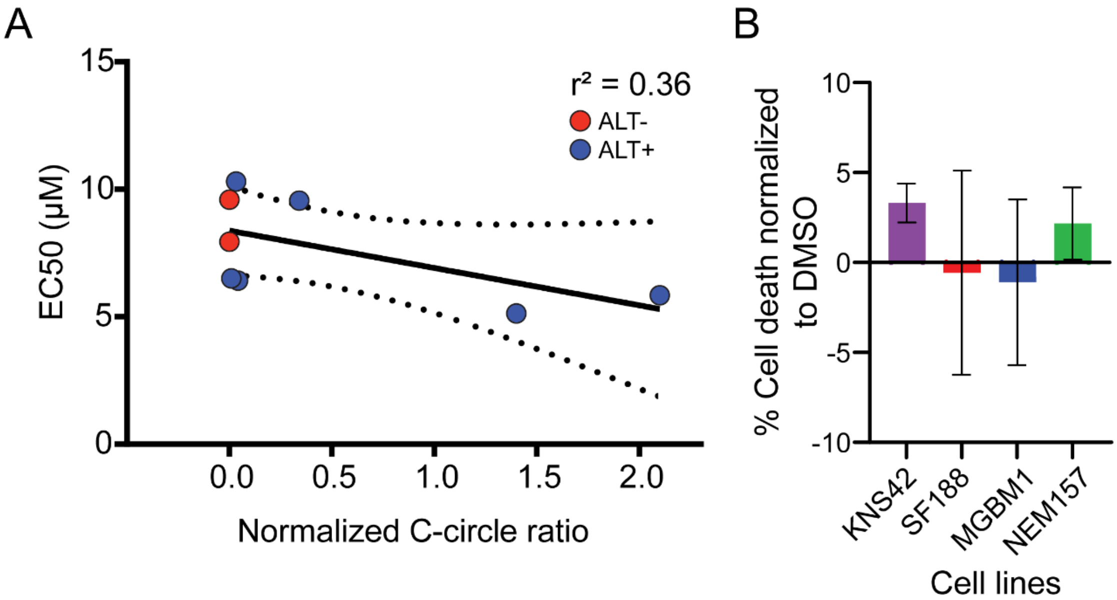


Figure 30. GSK343 lethality experiments and characterization. A) EC_{50} for cell survival in two ALT- cell lines (red) and 6 ALT+ cell lines (blue). The X-axis shows the normalized C-circle ratios for each cell line. The sample size consists of three technical replicates repeated three times. r^2 denotes the Pearson correlation. B) % cell death induction in four cell lines treated with 6 µL GSK343 for six days. Average of three technical replicates repeated three times.

5.5 Establishing imaging methods for H3.3S31p

To investigate the ALT-specific aberrant H3.3S31p, which consists of chromosome-wide spreading rather than the usual centromeric localization of H3.3S31p. I established a protocol for arresting cells in mitosis and image condensed chromosomes using antibody labeling and immunofluorescence. I included a pan-kinase inhibitor to function as a positive control for future H3.3S31p inhibitor testing, which potently diminished H3.3S31p signal (Figure 31A). Another aspect was to establish antibody stainings that could discern between aberrant and centromeric H3.3S31p spreading. This would be used to test inhibitors that could cause aberrant H3.3S31p in non-aberrantly spreaded cells or classify aberrant spreading in new cells. Therefore, the centromeric histone CENP-A was chosen and this signal largely overlapped with H3.3S31p in a non-spreaded cell line (Figure 31B).

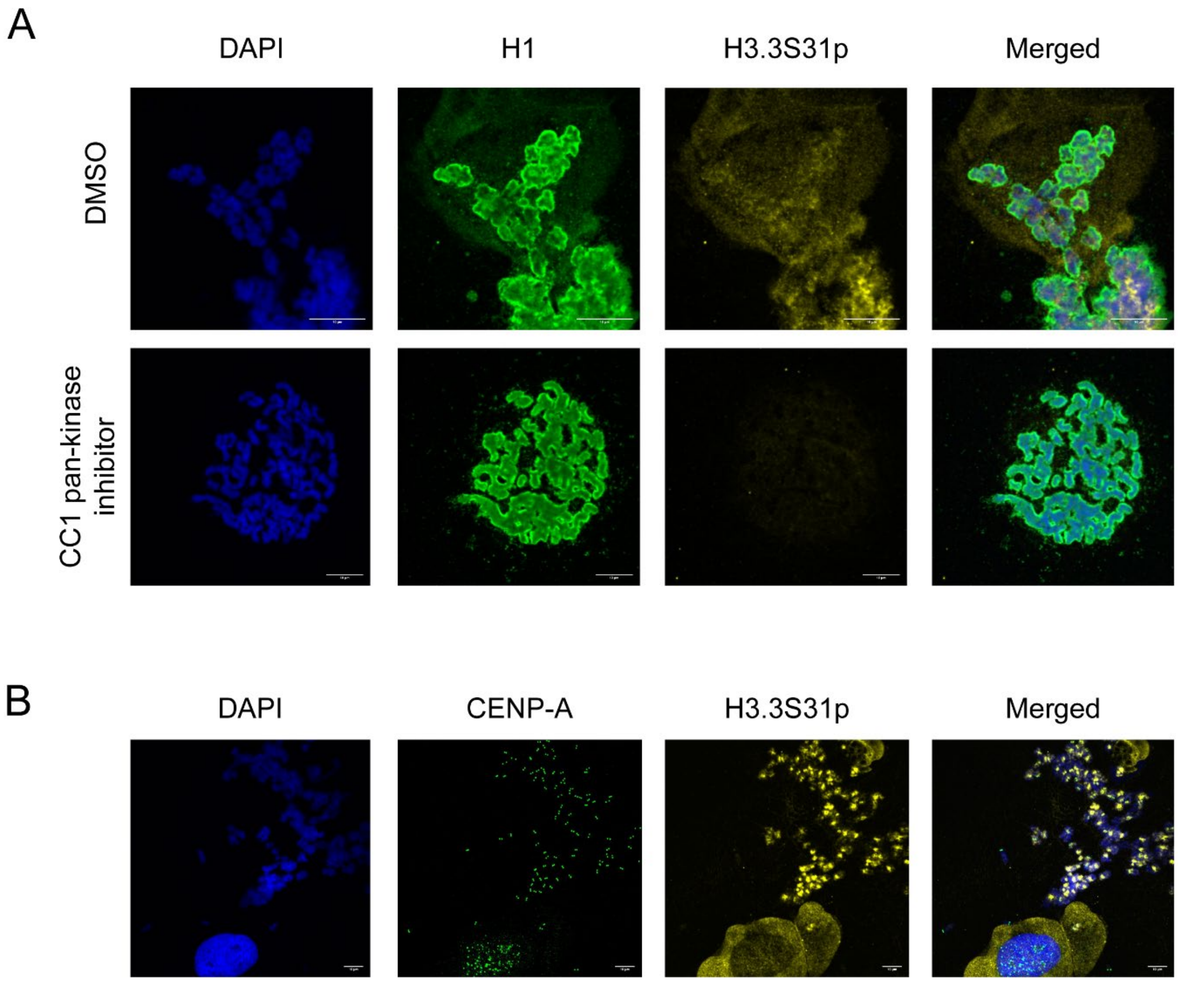


Figure 31. Metaphase spread optimizations. The white bar shows 1 μm length. (A) New protocol for MGBM1 cells using H1 as counterstaining, together with DAPI and H3.3S31p. Comparison between DMSO and CC1 pan-kinase inhibitor. (B) Testing CENP-A as counterstaining for centromeric H3.3S31p, using KNS42. N=1.

5.6 Aberrant H3.3S31p spreading in ALT cells may involve the mitotic kinase HASPIN

The kinase(s) responsible for depositing aberrant H3.3S31p has not been conclusively found to this date. A preliminary experiment indicated that inhibiting a mitotic kinase called HASPIN reduced H3.3S31p signal intensity (Figure 32B). In contrast to previous results studying the CHK1 kinase (64), I could not find a signal decrease when using a CHK1 inhibitor.

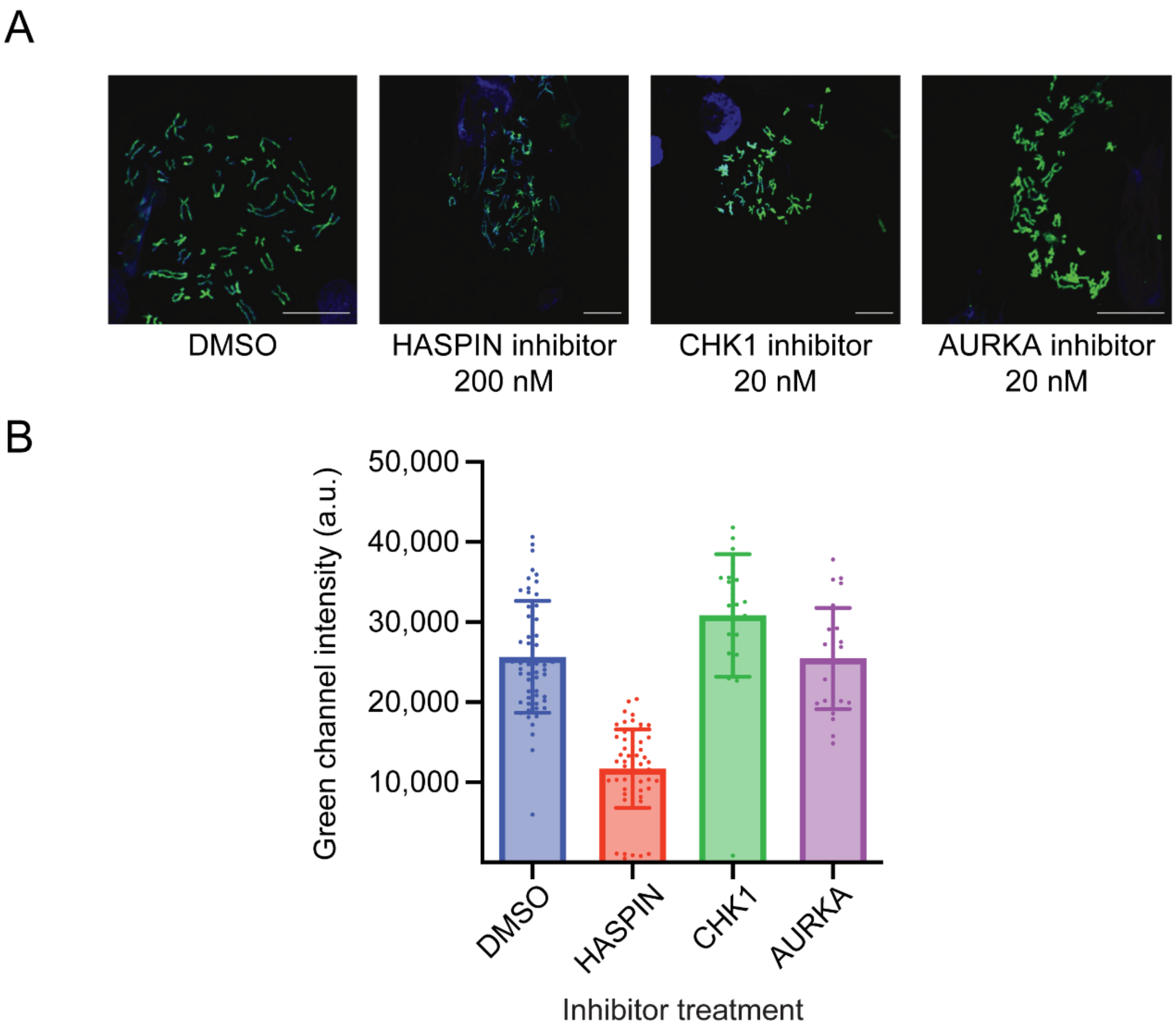


Figure 32. Metaphase spreads of the MGBM1 cell line treated with kinase inhibitors. (A) Inhibitors LDN-192960 (HASPIN), LY2603618 (CHK1), MLN8237 (AURKA), or DMSO were used. Merged images of blue fluorophore being DAPI and green fluorophore α-H3.3S31p Alexa Fluor 488. The white bar shows 1 µm length. B) Bar plot showing H3.3S31p intensity of segmented chromosomes when treated with DMSO, HASPIN, CHK1, or AURKA inhibitors. N=1.

Discussion

This thesis characterizes molecular features of the ALT pathway that depends on aberrant DNA repair and recombination, where many factors cooperate to amplify replicative stress that elongate telomeres via aberrant DNA repair. ALT can affect long-term survival in cancers and obtaining an omics signature would thus be valuable for patient stratification. Since ALT can also coexist with telomerase cancers (57), interventions targeting telomere elongation in cancer will require combinatorial treatment regimens that equally affect telomerase and ALT tumors. To get to this point, however, will require better knowledge of how ALT functions. In this thesis, I describe that ALT cancers are highly heterogeneous and yet have common features across tumor entities. I find that ALT cancers display distinct profiles concerning gene and miRNA expression and chromatin accessibility. I further characterize gene regulation networks that modulate the genetic environment in the presence of ALT. Using an integrative omics approach, I dissect tentative biomarkers and extract a multi-omics ALT signature. Thereafter, using a panel of chemical inhibitors on pediatric glioblastoma (pGBM) ALT cell lines, I find a relationship between ALT-activity and sensitivity to EZH2 and DNA methylation inhibitors. Lastly, by assaying different kinase inhibitors, I report initial results that point to HASPIN as a kinase responsible for aberrant H3.3S31p spreading during mitosis in ALT.

1 C-circle analysis of 687 mixed-lineage sarcomas show heterogeneous ALT activities

The C-circle assay has proven the most reliable assay for detecting ALT presence (66). Nevertheless, there appear to be some ALT cell lines that do not have high C-circle levels (Table 4). I find that out of 687 mixed-lineage sarcomas, 148 (22 %) were C-circle positive (Results section 1.2). The same percentage was obtained in another study that assayed ALT+ cancers in mixed-lineage sarcomas (220). The study in question used telomere-FISH to quantify telomere length, which is a more downstream biomarker for ALT. This would further indicate that C-circle amplification is an accurate way to capture ALT presence.

C-circles are quantitatively correlated with ALT activity. When using gamma-radiation to induce high ALT activity via telomeric double-strand breaks exogenously, C-circles are increased (221). This notion is also supported in other studies that modulate replication stress factors or induce DNA damage to increase ALT activity (119, 222-224). The large spread in C-circle ratios I found indicates that ALT cancers are heterogeneous in their activity. This finding is also not restricted to ALT sarcomas. A study conducted in 720 neuroblastomas saw a similar large spread in C-circle values (52). Taken together, my findings show that the C-circle assay accurately captures ALT presence and that ALT cancers have a spectrum of recombinational activity. The cause for the heterogeneity is not yet confirmed; however, it is possible that the large spread in ALT activity constitutes a spectrum of increased DNA damage and telomere recombination. Interestingly, new findings reveal that two general telomere elongation mechanisms exist in ALT and that only one of these mechanisms may contribute to C-circle formation (69). Thus, there is a possibility that some ALT cancers would remain undetected using the C-circle assay.

2 Modulating H3K27me3 may be related to ALT-activity irrespective of existing H3.3 mutations

With all the inferred epigenetic changes in ALT, it should come as no surprise that ALT cell lines show differential viabilities when treated with epigenetic drugs (see Results section 5.1. The trend towards decreased viability to the EZH2 inhibitor GSK343 in high-activity ALT samples (Figure 30A) would imply that ALT-activity is underpinned by epigenetic derepression irrespective of H3.3 mutations. Additionally, Figure 30B shows that the loss of viability resulted from cell lines not proliferating. A study on diffuse intrinsic pontine glioma (DIPG) cells discerned that GSK343 treatment resulted in a loss of viability in cancers with H3K27M mutations (219). Similarly, the authors found that GSK343 inhibited cell proliferation and saw via RNA-seq that this was due to the expression of the oncosuppressor $p16^{ink4a}$. $p16^{ink4a}$ is a common cancer mutation due to its senescence-inducing capabilities (225) and could explain the sudden stop in cell proliferation found in both this study as well as the cell death analysis that was performed (Figure 30B). It is known that DIPGs are enriched in ALT, which also contain H3K27M or H3.3K27M mutations (226). However, the authors did not check the presence of ALT in their tumor samples. Thus, further work is needed to fully characterize why ALT cells stop proliferating during PRC2 inhibition by assessing whether treated cells become senescent with β-galactosidase assays.

3 DNA methylation perturbation leads to differential survival connected to ALT-activity

The mode of action of 5-azacytidine's mechanism is not entirely understood, although its inhibitory effect on DNA methylation is well-established (227). Incorporating 5-azacytidine instead of cytosine residues trap DNA methyltransferases and leads to global DNA methylation reduction (227). The lack of DNA methylation can induce cytotoxicity either by modulating gene expression programs or inducing innate immunity from expressed TEs (228). Therefore, the increased survival in high-ALT cell lines would indicate that ALT-activity confers resistance to these processes. It could also be inferred that innate immunity is heavily disrupted in ALT and that expression of TEs does not yield inflammation-mediated cell death. Another possibility is that methylation levels at ALT subtelomeres are already low. It has been shown that subtelomeres are hypermethylated when telomerase is active (229). While the methylation status of ALT subtelomeres has not been assayed, it is possible that they are hypomethylated. Furthermore, telomere position over long distances (TPE-OLD) studies showed that short telomeres lead to DNA methylation patterns which allow TERT expression (230). As such, long telomeres, which are unique to ALT cancers, may by themselves change subtelomeric DNA methylation in a hitherto unknown fashion. Therefore, it could be so that the cells with high ALT activity have longer telomeres that sustain DNA methylation through 5-azacytidine-insensitive mechanisms.

4 Haspin may coordinate H3.3S31p deposition

As previously mentioned, H3.3S31p is usually a pericentric mitotic histone mark. In the majority of ALT cell lines, the mitotic spreading for H3.3S31p is chromosome-wide, and this aberrant spreading phenotype can be a good ALT marker (64). Furthermore, H3.3S31p drives enhancer activation in embryonal stem cells and might aid in resolving UV damage in DNA (115, 116). Despite this, it is unclear how H3.3S31p is involved in ALT, as it is aberrantly deposited in a chromosome-wide manner for many but not all ALT cell lines (64). Multiple studies have attempted to elucidate which kinase(s) are involved in its deposition. One study used a siRNA screening platform to investigate which kinase knockdowns led to a reduced H3.3S31p signal (231). They observed that the kinases AURKB and BUB1B, which drive mitotic checkpoint processes together, reduced the S31 phosphorylation greatly (232). However, it is worth mentioning that this study was performed on HEK293 cells, which are ALT- (233). Another study done in ALT+ cell lines showed that the DNA-repair kinase CHK1 is involved (64). However, the authors used inhibitor concentrations multiple orders of magnitude above the IC_{50} value. In the inhibitor study I performed, I chose values close to the IC_{50} values to avoid potential side effects. In my experiments, the only inhibitor that reduced the mitotic H3.3S31p signal was the HASPIN inhibitor. HASPIN has been described to phosphorylate threonine 3 in H3 (H3T3) in mitosis and has been implicated in proper chromosomal alignment (234). HASPIN activity also depends on phosphorylation by AURKB (235). Given the genetic instability present in ALT, it is conceivable that mitotic defects go together with this instability. Thus, findings from an earlier study that observed the function of H3.3S31p as a sensor for misaligned chromosomes could apply to aberrant H3.3S31p ALT cells (117). This model would suggest that H3.3S31p is a consequence of ALT-induced genetic instability, rather than it being involved in differentiation or gene expression pathways. Future studies should use AURKB inhibitors to verify its relationship to H3.3S31p. Additionally, the cause-or-consequence question for aberrant H3.3S31p should be investigated further as well. In this regard, it would be beneficial to induce aberrant H3.3S31p spreading in ALT cell lines with normal spreading patterns and test if this affects ALT activity. Another experiment to perform in the same cell line would be to induce chromosomal missegregation using microtubule inhibitors and observe whether aberrant H3.3S31p is formed.

5 Different omics approaches identify ALT specific features

C-circles are established biomarkers for ALT, but omics signatures for ALT are lacking. Since ALT can profoundly influence long-term survivability in cancers (53, 236), it is vital to investigate whether ALT presence can be linked to sequencing-based readouts. In this study, I performed RNA-seq of long and short RNAs as well as ATAC-seq on a heterogeneous tumor panel with varying mutations in ATRX, H3.3, and ALT activity with the approach depicted in Figure 33. One challenge is that tumor entity could confound the ALT signature, which needs further consideration in future validation experiments. This is evidenced by the eigenfactor correlation analysis I performed (Figure 12), where both PC1 and PC2 showed a high correlation to both TMM as well as tumor entity for most omics approaches.

Regarding the ALT signature found in the different omics, it appears that the most differential features are obtained from transcriptome, chromatin accessibility, and miRNA. This is also reflected when performing multi-omics factor analysis (MOFA) (see Results section 4.2). In this regard, I obtained the *NFATC2* and *BRSK2* genes that showed both differential expression in transcriptome and chromatin accessibility. *NFATC2* could also be confirmed through transcription factor (TF) analysis from upregulated ALT transcriptomic genes, indicating that its activation can drive the differentially expressed ALT genes. However, these genes would have to be functionally validated and observed in primary cancers. One extensive ALT omics study was performed in neuroblastoma cancers, where RNA-seq and proteomics were used to identify ALT-related proteins and genes (52). The authors did not observe a differential expression of my candidate hits *BRSK2* and *NFATC2.* Both of these genes may be down or upregulated via inflammation-related processes (see Discussion section 7), and one explanation could be that ALT neuroblastomas are less prone to inflammation. Similarly, a study used the TCGA cancer cohort to dissect ALT-related genes and found enrichment of GO-terms relating to reactive oxygen species, hypoxia, and angiogenesis in most ALT tumors (237). However, the TCGA cohort does not include pediatric cancers, and thus, neuroblastoma is not included in this analysis (238).

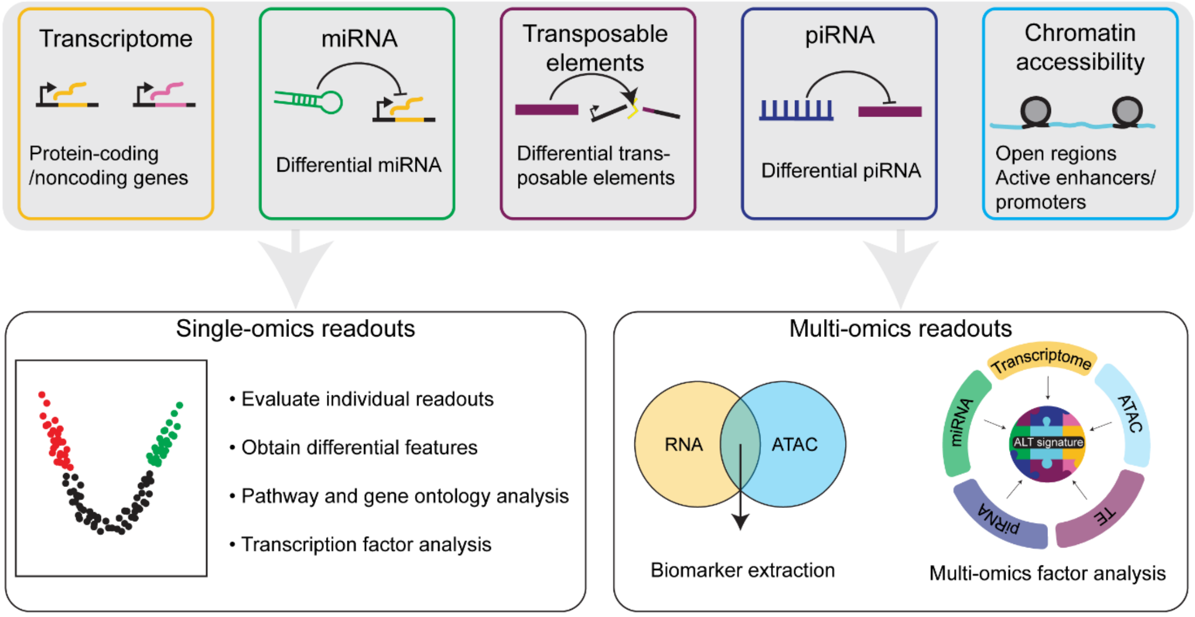


Figure 33. The different sequencing readouts when evaluating ALT omics signature

6 Role of SOX and RUNX transcription factors in ALT

ATAC-seq identified more than 5,000 differentially accessible peaks in ALT+ cell lines, where most peaks were found in non-coding genetic elements (Figure 21). RUNX TF family motifs were enriched while SOX motifs were depleted, indicating that ALT+ cell lines selectively utilize these developmental pathways. SOX and RUNX TFs are crucial in diverse differentiation processes and similarly have a wide range of chromatin targets (226, 227). These TFs were not found as differentially expressed genes in the transcriptome TF analysis. However, it is well established that post-translational modifications also regulate TF activity, and these TFs may thus be vital in inducing ALT-specific developmental states (208). It is also interesting that the transcriptome contained many developmentally related GO-terms, which could be regulated by the SOX and RUNX TFs (Table 5 and Table 6). Another open question is why these specific TF families are up or downregulated. One potential mechanism could be miRNAs. It was shown that SOX4, one of the downregulated motifs in my TF analysis, can be downregulated or have its activity inhibited by miR-138 and miR-142 (239, 240). These miRNAs were upregulated in my differentially expressed miRNA analysis (Supplemental Table 3). Similarly, for RUNX, one of the upregulated differentially expressed miRNA miR-218 increased RUNX2 expression in osteoblasts (241). The reason for the RUNX TFs being upregulated may also be due to higher activity of the AP-1 family, as well as STATs and various other immune-related TFs (242-244) (Figure 15). Lastly, RUNX TFs may induce specific pathways not captured by chromatin accessibility or transcriptome TF analysis. The pathway analysis revealed that RUNX1 induced gene expression pathways involving estrogen receptor activation, WNT signaling, and differentiation of myeloid cells. The pathway analysis of down-weighted genes (PADOG) improves gene set analysis significantly and could explain why similar GO-terms were not enriched (201).

7 AP-1 and immune-related transcription factors are indicative of inflammation in ALT+ cell lines

The highest enrichment of TF motifs found in the chromatin accessibility analysis was for the AP-1 family. These TFs can heterodimerize in multiple combinations to result in different transcriptional profiles (209). Notably, AP-1 TFs can be activated in response to DNA damage and inflammation (210, 245). These stressors may be a common theme in ALT cancers and could cause specific upregulation of AP-1 TFs. Interestingly, in an ALT induction model, the authors saw increased inflammatory markers (such as TGFβ) but reduced AP-1 TFs (246). The authors do not go into great detail explaining this pathway. However, it is possible that in transitioning to become ALT+, the cancer cells modulate downstream pathways of inflammation, such as AP-1, to avoid apoptosis. This would also indicate that cell lines that have ALT rely on high AP-1 activity at a later stage. However, an open question is how the ALT cell lines disrupt inflammation so as not to show signs of, e.g., TGFβ signaling in GO-term analysis.

It was recently shown that C-circles induce an innate immune response and that ALT cancers disrupt DNA sensing mechanisms to avoid apoptosis (216). This could explain why modulating key components of the DNA sensing pathway did not result in differential viability (see Results section 5.1), as these pathways are simply quiescent. Nevertheless, it would also make sense that the ALT cancer utilizes this heightened state of inflammation to drive tumor growth. This process has been described in many cancer-related features, such as tumor microenvironment and cancer proliferation (247). Upregulated genes show many enriched immune-related transcription factors for ALT cell lines, such as STATs and NFATs (Figure 15B). This indicates that existing inflammatory processes influence ALT transcriptome. The identification of *NFATC2* from transcriptomic data further supports this (see Results section 3.3). *NFATC2* is a transcription factor generally expressed in resting T-cells (248), yet it was highly upregulated in ALT+ cell lines (Figure 14) and amongst the top TF motif in upregulated ALT+ genes (Figure 15B).

C-circles and chromosomal instability may account for some of the inflammatory signaling. In addition, the multi-omics factor analysis with the MOFA software revealed that ALT+ cell lines had a reduced response to oxidative stress (Figure 27B), which could in turn lead to an increase in reactive oxygen species mediated damage. Oxidative stress is a significant source of inflammation (249), and telomeres in particular are susceptible to oxidative damage (250). Additionally, GO-terms related to negative regulation of neuron death and cellular protein assembly may further indicate the presence of oxidative stress. Neurons are susceptible to reactive oxygen species and mitochondrial dysfunction (213), and so is the protein assembly complex (251). It was shown within the TCGA cancer cohort that GO-terms relating to reactive oxygen species were upregulated in most ALT tumors (237).

Together with findings that mitochondria may be dysfunctional in ALT (128), these studies and my own study suggest that oxidative stress and mitochondria are connected to the ALT pathway. However, it can additionally be seen in the MOFA signature that oxidative stress damage is actively promoted in ALT and could possibly be a critical route for cells to induce DNA damage and drive telomeric recombination. It can also be speculated that ER stress is induced in ALT due to reactive oxygen species. This may also be related to one of the top downregulated genes in ALT+, namely *BRSK2*. It was shown that *BRSK2* mediates ER-stress-induced apoptosis and that its expression is downregulated in cancer cell lines to escape cell death (252).

The diverse set of findings from the omics analysis may be related to the heterogeneity in ALT activity described in Results section 1.2. The large spread in ALT activity can be underpinned by different degrees of DNA damage which facilitate DNA recombination. Inflammation would go hand-in-hand with increased levels of genetic instability (253), and it is known that inflammation can be modulated into pro-survival processes in cancer (254). This can be confirmed experimentally by inducing cell death, for example, whether apoptosis is induced to the same degree in ALT+ cell lines. Furthermore, it is also possible that a constant state of inflammation can lead to ALT tumors being more infiltrated by immune cells (255). Additionally, as ALT cancers have a heterogeneous level of activity, it can be postulated that ALT cancers with high activity can lead to increased levels of immune cell infiltration. This could be utilized in immune checkpoint blockades, such as CTLA-4 and PD-1/PDL-1 therapeutics (256).

8 ALT transcriptomic signature indicates modulation of signaling pathways

While the GO-terms related to differentially expressed ALT genes were quite broad and included many developmental pathways, some enrichment of neuronal pathways was apparent when doing pathway analysis (Table 8). It pointed to activin/inhibin as well as Neuropilin signaling. Activin ligands are a member of the TGF-β superfamily of ligands (257). The activin ligands bind to activin receptors, which induce gene expression changes via SMAD and MAPK pathways and are vital in cell proliferation (258-260). Activin signaling has also been implicated in many inflammatory diseases and cancers (261, 262), and thus its activation could be due to increased inflammation in ALT. The activin signaling pathway is negated by the inhibin ligands and follistatin, which compete for the same receptors but result in non-productive signaling (263, 264). One of the top upregulated ALT+ genes is *INHBA*, an inhibin ligand. Perhaps the ongoing inflammation in ALT+ cell lines leads to activin signaling, which is inhibited by upregulating *INHBA*. Thus, it would be worthwhile to investigate whether activin gene expression is detrimental to ALT, for example, by performing knockouts of *INHBA*.

Semaphorin pathways and three different VEGF pathways indicate that ALT+ cell lines are distinct with respect to angiogenetic pathways (265). Neuropilin is a neuronal membrane receptor that binds to semaphorin and VEGF ligands and has diverse functions in cancer (266, 267). The ALT cell lines appear to favor the neuropilin signaling axis while inhibiting the canonical VEGF/VEGFR axis. Another possible gene involved in angiogenesis is *PTN*, which was the most upregulated ALT gene. The PTN protein is a growth factor expressed in many cancers (268-270). PTN is a potent inducer of angiogenesis, with potential mechanisms including *VEGF* upregulation or binding to the receptors $\alpha_v\beta_3$ integrin and PTPRZ1 (271-273). However, the reason why the ALT cell lines require upregulation of angiogenesis is unclear. It is known that hypoxia and inflammation can upregulate angiogenesis (274), and both of these processes may occur to a higher degree in ALT cells. Hypoxia can be crucial in producing reactive oxygen species (275), and together with the reduction in oxidative stress response found in the multi-omics ALT signature, hypoxia could be yet another cause of oxidative stress. Within the TCGA study of ALT cancers, hypoxia and angiogenesis were also observed (238). Nevertheless, the lack of hypoxia-related findings in the sequencing results may rule this out. Hypoxia-related GO-terms should be present in the GO-term analysis, as well as finding the hypoxia TF HIF1α in the TF motif analysis. A possibility is that ALT cells utilize different pathways for their maintenance, and to confirm this, sequencing studies including ALT cell lines from one of the tumor entities that show a hypoxia phenotype would need to be performed.

9 Small RNA-seq identifies many telomere-related miRNAs

Of the 52 differentially expressed miRNAs identified in ALT cell lines (see Results section 3.8), 17 had the GO-term “negative regulation of telomere elongation”. However, it is worth noting that miRNAs may have thousands of potential binding targets. miRNA bind and repress mRNA transcripts via complementary sequences (276), and target prediction databases are mainly concerned with this mechanism (277). The GO-terms were calculated by performing enrichment analysis of the miRNA targets. Yet, with thousands of putative mRNA targets, each miRNA can have many different GO-terms associated with it. Additionally, miRNAs can themselves be regulated via other miRNA or via lncRNA that act as miRNA "sponges" (278). Therefore, miRNA functions can be heavily dependent on a specific context that must be considered.

A newly published paper investigated miRNA in ALT+ cell lines using microarrays (279). Their top hit was miR-708, the second-highest miRNA in terms of adjusted *p*-value from my differentially expressed miRNA analysis. The authors of the study characterized this miRNA, whereby its overexpression led to a loss of invasion and angiogenesis via repression of the TF CARF. The authors observed that this miRNA suppressed proliferation in ALT- cells but promoted it in ALT+ cell lines. This was believed to occur via regulation of BRCA1 and MRE11, which are key components of homologous recombination (HR) (280). The genes for these proteins, together with CARF, were not found in my differentially expressed genes or the TF analysis for CARF. This could be due to limitations in the target prediction, which, as previously stated, are computationally calculated and are primarily focused on miRNA-mRNA interactions. Another miRNA that was studied in the context of telomerase was miR-512. It was found downregulated in ALT- head and neck cancers, most likely due to degrading TERT (281). This suggests that ALT cancers require active repression of TERT, in this case via the miR-512. Altogether, the differential miRNAs found in my sequencing analysis tentatively show a role of these regulatory moieties in ALT. While experimental evidence is lacking, they may influence gene expression pathways that sustain ALT. One example could be to downregulate telomerase and, by doing so, induce ALT. Another would be to sustain cell proliferation by modulating HR proteins. As discussed in the sections above, some miRNAs regulate TF activity and expression and may explain how the specific TF families are induced or suppressed in ALT.

10 Transposable elements and piRNA involvement in ALT

Transposable elements (TE) are viral remnants from ancient infections, and subtypes of TEs can still be transcribed into active proteins (282). This is exemplified by the long-interspersed element-1 (LINE-1) families, which can be retrotransposed and integrated into genes *de novo (283)*. From an evolutionary perspective, LINE-1 has been central in forming pseudogenes by inserting and duplicating itself with functional genes (284), and in recent years LINE-1 expression has been identified as a hallmark of human cancers (283, 285). LINE-1 insertions within coding genes can disrupt gene expression or regulatory elements to stifle oncosuppressor expression further (260). A counterbalance to TEs are the piRNAs, which like miRNA, degrade transcripts that have sequence complementary. The piRNAs also induce epigenetic silencing of the active TE locus using DNA and histone methylation (286). In recent years it was also shown that piRNAs could mediate protein-coding gene expression either via mRNA degradation or by inducing epigenetic silencing (287-289). My genomic analysis found no differentially expressed TE families and a handful of piRNAs (see Results section 3.9). The first finding is surprising, as the additional genetic instability incurred by expressed TEs could theoretically further drive ALT. Additionally, the increased chromatin accessibility in ALT could also lead to TE expression. Three reasons for this are possible. (i) TEs are already transcriptionally silent due to piRNAs and other epigenetic repression mechanisms. (ii) The strategy for generating TE count tables requires further optimization. (iii) There are no differences in TE transcription between ALT+ and ALT- cell lines.

With respect to the first point, I found 14 differentially expressed piRNAs (Figure 19B), which would indicate that TEs are already being repressed. This is further supported by two upregulated genes involved in piRNA-mediated repression found when doing the TelNet overlap (Table 7), namely *TDRD6* and *TDRKH*. These are Tudor domain-containing proteins, which are crucial in synthesizing piRNAs and repressing TEs (290). However, due to lacking target prediction databases, the targets for the differentially expressed piRNAs would have to be investigated. This can be addressed by downregulating piRNA expression or inducing TE expression. To address the second question, the TE computational analysis is performed either from individual transcripts (e.g., L1PB2_dup77), followed by the family (e.g., LINE-1), and then by the class (e.g., LINE). Individual transcripts are not well suited for assaying TE expression, as one TE transcript maps to many different individual transcripts (291). Therefore, the best practice in identifying TEs is to look at the family or class expression. To address the third question, there is also the chance that TEs are equally repressed in ALT-/ALT+ cell lines and that there is no ALT-specific mechanism by which TEs are derepressed. Although ChIP-seq or ATAC-seq would be employed to observe whether TE sites gain active genetic marks or chromatin accessibility, TEs are often found at so-called blacklisted sites (292), which are prone to artifacts when using sequencing methods that are based on chromatin shearing (ChIP-seq) or transposase (ATAC-seq). Regardless, it is interesting that the TEs and piRNAs are both equally present in the multi-omics ALT signature (Figure 27A). However, their contribution to the ALT-signature was minor, and altogether this would indicate that TE and piRNA expression are not good predictors of ALT.

Conclusion and Future Directions

This thesis studied how specific transcription and epigenetic programs could translate into functional downstream effects crucial for ALT cancer cells, as summarized in the scheme depicted in Figure 34. The TF analysis by ATAC-seq and RNA-seq identified an increased activity of RUNX and AP-1 family members in ALT. AP-1 may be activated by inflammation, which in turn can drive RUNX activity (210, 244). Inflammation can also be the cause of immune-related TFs found in upregulated genes. Downregulated ATAC peaks are enriched in TF motifs from the SOX family, where one specific mode of repression may be the increase of miR-138/142, which modulates SOX4 activity (239). Additionally, downregulated genes are enriched in the EGR-1/2, c-Myc and Hic-1 TFs. The gene expression layer is upregulated by the immune TFs, and RUNX1 as evidenced from pathway analysis. The down- and upregulated genes then converge into developmental processes, as identified through GO-term analysis. Specific developmental pathways may also be upregulated due to hypomethylation and/or low H3K27me3. Additionally, the MOFA signatures relating to oxidative stress may explain the downregulated gene *BRSK2* (252). The cellular protein assembly is negatively weighted, as is the negative regulation of neuron death. The latter signature may be related to neuronal signaling, which in turn is modulated by the developmental processes. Additionally, this layer contains miR-512, which can downregulate *TERT* expression (281). The functional downstream effects of the developmental processes can be connected to the top upregulated ALT genes *PTN* and *INHBA*. PTN has a role in angiogenesis, where it interacts with VEGF (271). As inferred from the pathway analysis, VEGF may induce signaling via neuropilin and semaphorin receptors to cause angiogenesis. *INHBA,* on the other hand, suppresses activin signaling and its downstream targets in gene expression (263). Lastly, the miR-708 upregulation might enhance proliferation of ALT cells (279).

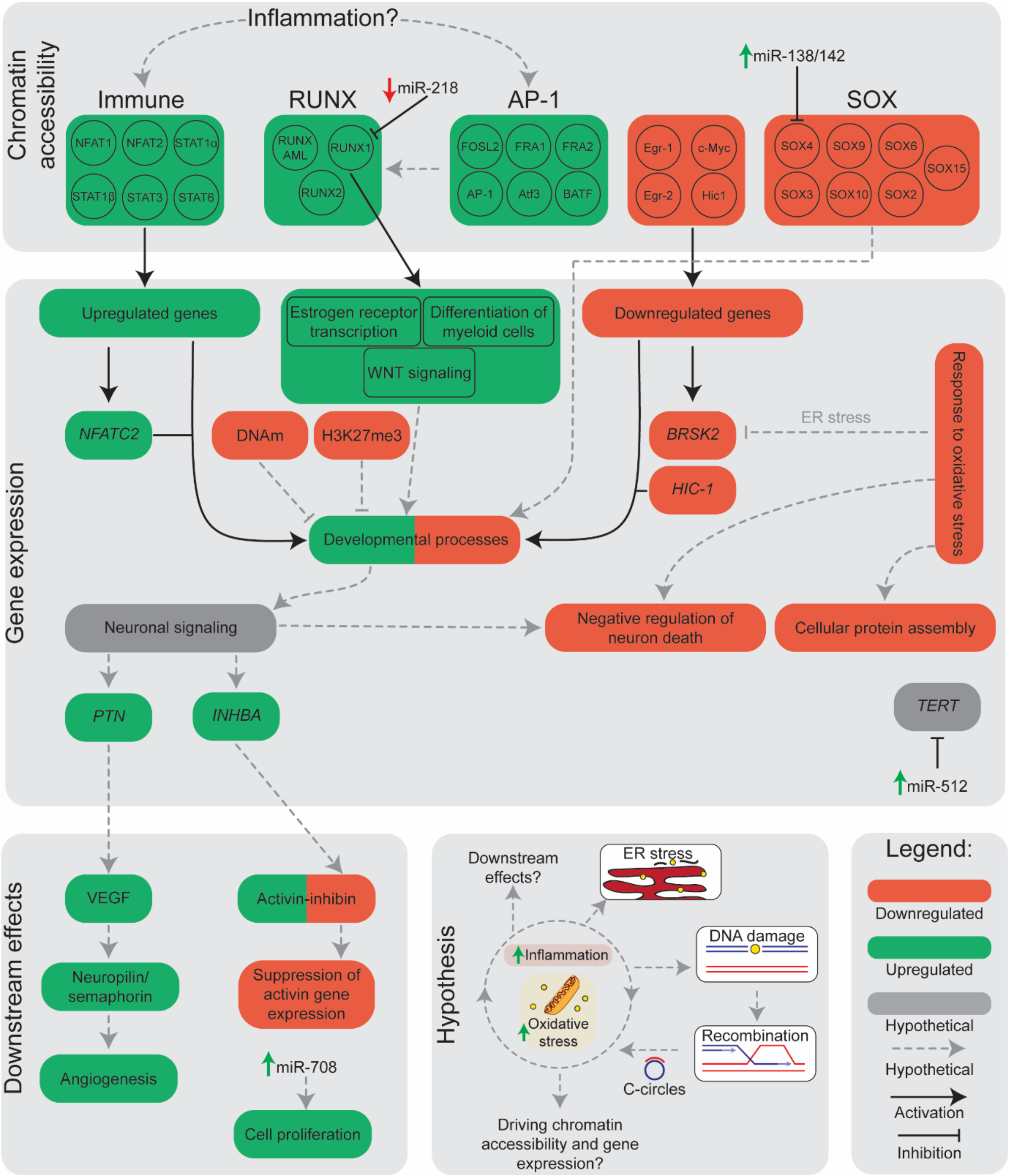


Figure 34. Layers of chromatin regulation and gene expression involved in ALT.

In summary, the multi-omics analysis revealed novel ALT features, which can guide future research. One crucial question from this analysis is emerging: Does high ALT activity also elicits more inflammation, and is oxidative stress signaling involved in this process? Using

ALT cancers with known C-circle levels and investigating the presence of 8-oxoguanine in chromatin could further our knowledge in this process. Additionally, performing proteomics or immune cytokine ELISA on these cancers would be beneficial in answering whether inflammation is induced by ALT. This could potentially result in ALT cancers with high activity being more susceptible to therapeutics targeting inflammation, such as drugs targeting immune checkpoints. Applying multi-omics ALT signatures as done here could be crucial in detecting the presence of ALT in tumor sequencing data, but this signature will have to be validated against primary tumor samples. This would have to be performed in tumors with known C-circle levels and ALT activity and then use the omics signature to predict TMM status. An omics-derived ALT signature would present a substantial advancement for predicting TMM status in a clinical context since telomere maintenance is a key cancer hallmark.

References

1. De Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005;19(18):2100-10.
2. Levy MZ, Allsopp RC, Futcher AB, Greider CW, Harley CB. Telomere end-replication problem and cell aging. J Mol Biol. 1992;225(4):951-60.
3. Huffman KE, Levene SD, Tesmer VM, Shay JW, Wright WE. Telomere shortening is proportional to the size of the G-rich telomeric 3’-overhang. J Biol Chem. 2000;275(26):19719-22.
4. Hayashi MT, Cesare AJ, Rivera T, Karlseder J. Cell death during crisis is mediated by mitotic telomere deprotection. Nature. 2015;522(7557):492-6.
5. Wright WE, Piatyszek MA, Rainey WE, Byrd W, Shay JW. Telomerase activity in human germline and embryonic tissues and cells. Dev Genet. 1996;18(2):173-9.
6. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. Eur J Cancer. 1997;33(5):787-91.
7. Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA, Reddel RR. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nat Med. 1997;3(11):1271-4.
8. Zhang JM, Zou L. Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. Cell Biosci. 2020;10(1):30.
9. Okuda K, Bardeguez A, Gardner JP, Rodriguez P, Ganesh V, Kimura M, et al. Telomere length in the newborn. Pediatr Res. 2002;52(3):377-81.
10. Friedrich U, Griese E, Schwab M, Fritz P, Thon K, Klotz U. Telomere length in different tissues of elderly patients. Mech Ageing Dev. 2000;119(3):89-99.
11. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell. 2010;40(2):179-204.
12. Verdun RE, Karlseder J. Replication and protection of telomeres. Nature. 2007;447(7147):924-31.
13. de Lange T. T-loops and the origin of telomeres. Nat Rev Mol Cell Biol. 2004;5(4):323-9.
14. Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell. 2009;138(1):90-103.
15. Opresko PL, Mason PA, Podell ER, Lei M, Hickson ID, Cech TR, et al. POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. J Biol Chem. 2005;280(37):32069-80.
16. Wu W, Rokutanda N, Takeuchi J, Lai Y, Maruyama R, Togashi Y, et al. HERC2 Facilitates BLM and WRN Helicase Complex Interaction with RPA to Suppress G-Quadruplex DNA. Cancer Res. 2018;78(22):6371-85.
17. Wang Y, Yang J, Wild AT, Wu WH, Shah R, Danussi C, et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nature communications. 2019;10(1):1-14.
18. Pfeiffer V, Lingner J. Replication of telomeres and the regulation of telomerase. Cold Spring Harb Perspect Biol. 2013;5(5):a010405.
19. Fouquerel E, Barnes RP, Uttam S, Watkins SC, Bruchez MP, Opresko PL. Targeted and persistent 8-oxoguanine base damage at telomeres promotes telomere loss and crisis. Mol Cell. 2019;75(1):117-30. e6.

20. Cicconi A, Chang S. Shelterin and the replisome: at the intersection of telomere repair and replication. Curr Opin Genet Dev. 2020;60:77-84.

21. Marcand S. How do telomeres and NHEJ coexist? Mol Cell Oncol. 2014;1(3):e963438.

22. Jia P, Her C, Chai W. DNA excision repair at telomeres. DNA Repair. 2015;36:137-45.

23. Hiyama E, Hiyama K. Telomere and telomerase in stem cells. Br J Cancer. 2007;96(7):1020-4.

24. Hockemeyer D, Collins K. Control of telomerase action at human telomeres. Nat Struct Mol. 2015;22(11):848-52.

25. Takubo K, Nakamura K, Izumiyama N, Furugori E, Sawabe M, Arai T, et al. Telomere shortening with aging in human liver. J Gerontol A Biol Sci Med Sci. 2000;55(11):B533-6.

26. Takubo K, Nakamura K, Izumiyama N, Sawabe M, Arai T, Esaki Y, et al. Telomere shortening with aging in human esophageal mucosa. Age (Omaha). 1999;22(3):95-9.

27. Prochazkova M, Chavez MG, Prochazka J, Felfy H, Mushegyan V, Klein OD. Embryonic versus adult stem cells. Stem Cell Biology and Tissue Engineering in Dental Sciences: Elsevier; 2015. p. 249-62.

28. Greider CW. Telomerase activity, cell proliferation, and cancer. Proc Natl Acad Sci U S A. 1998;95(1):90-2.

29. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194-217.

30. Blasco MA. Telomere length, stem cells and aging. Nat Chem Biol. 2007;3(10):640-9.

31. Blasco MA. The epigenetic regulation of mammalian telomeres. Nat Rev Genet. 2007;8(4):299-309.

32. Lejnine S, Makarov VL, Langmore JP. Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. Proc Natl Acad Sci U S A. 1995;92(6):2393-7.

33. Bandaria JN, Qin P, Berk V, Chu S, Yildiz A. Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. Cell. 2016;164(4):735-46.

34. Grewal SI, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8(1):35-46.

35. Riethman H, Ambrosini A, Paul S. Human subtelomere structure and variation. Chromosome Res. 2005;13(5):505-15.

36. Cubiles MD, Barroso S, Vaquero-Sedas MI, Enguix A, Aguilera A, Vega-Palas MA. Epigenetic features of human telomeres. Nucleic Acids Res. 2018;46(5):2347-55.

37. Marion RM, Strati K, Li H, Tejera A, Schoeftner S, Ortega S, et al. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. Cell Stem Cell. 2009;4(2):141-54.

38. Lou Z, Wei J, Riethman H, Baur JA, Voglauer R, Shay JW, et al. Telomere length regulates ISG15 expression in human cells. Aging (Albany NY). 2009;1(7):608-21.

39. Robin JD, Ludlow AT, Batten K, Magdinier F, Stadler G, Wagner KR, et al. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. Genes Dev. 2014;28(22):2464-76.

40. Kim W, Shay JW. Long-range telomere regulation of gene expression: Telomere looping and telomere position effect over long distances (TPE-OLD). Differentiation. 2018;99:1-9.

41. Toubiana S, Selig S. Human subtelomeric DNA methylation: regulation and roles in telomere function. Curr Opin Genet Dev. 2020;60:9-16.

42. Yehezkel S, Segev Y, Viegas-Pequignot E, Skorecki K, Selig S. Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. Hum Mol Genet. 2008;17(18):2776-89.

43. Deng Z, Campbell AE, Lieberman PM. TERRA, CpG methylation and telomere heterochromatin: lessons from ICF syndrome cells. Cell Cycle. 2010;9(1):69-74.

44. Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat–containing RNA and RNA surveillance factors at mammalian chromosome ends. Science. 2007;318(5851):798-801.

45. Rippe K, Luke B. TERRA and the state of the telomere. Nat Struct Mol Biol. 2015;22(11):853-8.

46. Pan X, Chen Y, Biju B, Ahmed N, Kong J, Goldenberg M, et al. FANCM suppresses DNA replication stress at ALT telomeres by disrupting TERRA R-loops. Sci Rep. 2019;9(1):1-14.

47. Oliva-Rico D, Herrera LA. Regulated expression of the lncRNA TERRA and its impact on telomere biology. Mech Ageing Dev. 2017;167:16-23.

48. Dilley RL, Greenberg RA. ALTernative Telomere Maintenance and Cancer. Trends Cancer. 2015;1(2):145-56.

49. Dagg RA, Pickett HA, Neumann AA, Napier CE, Henson JD, Teber ET, et al. Extensive Proliferation of Human Cancer Cells with Ever-Shorter Telomeres. Cell Rep. 2017;19(12):2544-56.

50. Brodeur GM. Spontaneous regression of neuroblastoma. Cell Tissue Res. 2018;372(2):277-86.

51. Brodeur GM, Bagatell R. Mechanisms of neuroblastoma regression. Nat Rev Clin Oncol. 2014;11(12):704-13.

52. Hartlieb SA, Sieverling L, Nadler-Holly M, Ziehm M, Toprak UH, Herrmann C, et al. Alternative lengthening of telomeres in childhood neuroblastoma from genome to proteome. Nat Commun. 2021;12(1):1269.

53. Hakin-Smith V, Jellinek DA, Levy D, Carroll T, Teo M, Timperley WR, et al. Alternative lengthening of telomeres and survival in patients with glioblastoma multiforme. Lancet. 2003;361(9360):836-8.

54. Reddel RR, Bryan TM, Colgin LM, Perrem KT, Yeager TR. Alternative lengthening of telomeres in human cells. Radiat Res. 2001;155(1 Pt 2):194-200.

55. ClinicalTrials.gov.

56. Bajaj S, Kumar MS, Peters GJ, Mayur YC. Targeting telomerase for its advent in cancer therapeutics. Med Res Rev. 2020;40(5):1871-919.

57. Perrem K, Colgin LM, Neumann AA, Yeager TR, Reddel RR. Coexistence of alternative lengthening of telomeres and telomerase in hTERT-transfected GM847 cells. Mol Cell Biol. 2001;21(12):3862-75.

58. Cerone MA, Londono-Vallejo JA, Bacchetti S. Telomere maintenance by telomerase and by recombination can coexist in human cells. Hum Mol Genet. 2001;10(18):1945-52.

59. De Vitis M, Berardinelli F, Sgura A. Telomere Length Maintenance in Cancer: At the Crossroad between Telomerase and Alternative Lengthening of Telomeres (ALT). Int J Mol Sci. 2018;19(2):606.

60. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40(6):463-71.

61. Zhuang XY, Yao YG. Mitochondrial dysfunction and nuclear-mitochondrial shuttling of TERT are involved in cell proliferation arrest induced by G-quadruplex ligands. FEBS Lett. 2013;587(11):1656-62.

62. Temime-Smaali N, Guittat L, Sidibe A, Shin-ya K, Trentesaux C, Riou JF. The G-quadruplex ligand telomestatin impairs binding of topoisomerase IIIalpha to G-quadruplex-forming oligonucleotides and uncaps telomeres in ALT cells. PLoS One. 2009;4(9):e6919.

63. Sobinoff AP, Pickett HA. Alternative Lengthening of Telomeres: DNA Repair Pathways Converge. Trends Genet. 2017;33(12):921-32.

64. Chang FT, Chan FL, JD RM, Udugama M, Mayne L, Collas P, et al. CHK1-driven histone H3.3 serine 31 phosphorylation is important for chromatin maintenance and cell survival in human ALT cancer cells. Nucleic Acids Res. 2015;43(5):2603-14.

65. Chung I, Osterwald S, Deeg KI, Rippe K. PML body meets telomere: the beginning of an ALTernate ending? Nucleus. 2012;3(3):263-75.

66. Henson JD, Cao Y, Huschtscha LI, Chang AC, Au AY, Pickett HA, et al. DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. Nat Biotechnol. 2009;27(12):1181-5.

67. Wu Y, Brosh RM, Jr. G-quadruplex nucleic acids and human disease. FEBS J. 2010;277(17):3470-88.

68. Henson JD, Neumann AA, Yeager TR, Reddel RR. Alternative lengthening of telomeres in mammalian cells. Oncogene. 2002;21(4):598-610.

69. Zhang JM, Yadav T, Ouyang J, Lan L, Zou L. Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. Cell Rep. 2019;26(4):955-68 e3.

70. Min J, Wright WE, Shay JW. Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. Genes Dev. 2019;33(13-14):814-27.

71. Mender I, Shay JW. Telomere Dysfunction Induced Foci (TIF) Analysis. Bio Protoc. 2015;5(22).

72. Londoño-Vallejo JA, Der-Sarkissian H, Cazes L, Bacchetti S, Reddel RR. Alternative lengthening of telomeres is characterized by high rates of telomeric exchange. Cancer Research. 2004;64(7):2324-7.

73. Varley H, Pickett HA, Foxon JL, Reddel RR, Royle NJ. Molecular characterization of inter-telomere and intra-telomere mutations in human ALT cells. Nat Genet. 2002;30(3):301-5.

74. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, et al. Break-induced replication is highly inaccurate. PLoS Biol. 2011;9(2):e1000594.

75. Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, et al. Altered telomeres in tumors with ATRX and DAXX mutations. Science. 2011;333(6041):425.

76. Yuen BT, Knoepfler PS. Histone H3.3 mutations: a variant path to cancer. Cancer Cell. 2013;24(5):567-74.

77. Feuerbach L, Sieverling L, Deeg KI, Ginsbach P, Hutter B, Buchhalter I, et al. TelomereHunter–in silico estimation of telomere content and composition from cancer genomes. BMC Bioinform. 2019;20(1):1-11.

78. Conomos D, Reddel RR, Pickett HA. NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. Nat Struct Mol Biol. 2014;21(9):760-70.

79. Kramara J, Osia B, Malkova A. Break-induced replication: the where, the why, and the how. Trends Genet. 2018;34(7):518-31.

80. Flynn RL, Heaphy CM. Surviving Telomere Attrition with the MiDAS Touch. Trends Genet. 2019;35(11):783-5.

81. Lezaja A, Altmeyer M. Dealing with DNA lesions: When one cell cycle is not enough. Curr Opin Cell Biol. 2021;70:27-36.

82. Zhang J-M, Genois M-M, Ouyang J, Lan L, Zou L. Alternative lengthening of telomeres is a self-perpetuating process in ALT-associated PML bodies. Mol Cell. 2021;81(5):1027-42. e4.

83. Mazzucco G, Huda A, Galli M, Piccini D, Giannattasio M, Pessina F, et al. Telomere damage induces internal loops that generate telomeric circles. Nat Commun. 2020;11(1):1-11.

84. Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. ATRX and DAXX: Mechanisms and Mutations. Cold Spring Harb Perspect Med. 2017;7(3):a026567.

85. Voon HP, Wong LH. New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone. Nucleic Acids Res. 2016;44(4):1496-501.

86. Watson LA, Solomon LA, Li JR, Jiang Y, Edwards M, Shin-ya K, et al. Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. J Clin Invest. 2013;123(5):2049-63.

87. Hu Y, Shi G, Zhang L, Li F, Jiang Y, Jiang S, et al. Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. Sci Rep. 2016;6(1):32280.

88. Wong LH, McGhie JD, Sim M, Anderson MA, Ahn S, Hannan RD, et al. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res. 2010;20(3):351-60.

89. Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ. ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. Hum Mol Genet. 1996;5(12):1899-907.

90. Gibbons RJ, Pellagatti A, Garrick D, Wood WG, Malik N, Ayyub H, et al. Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS). Nat Genet. 2003;34(4):446-9.

91. Dhayalan A, Tamas R, Bock I, Tattermusch A, Dimitrova E, Kudithipudi S, et al. The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. Hum Mol Genet. 2011;20(11):2195-203.

92. Li Y, Syed J, Suzuki Y, Asamitsu S, Shioda N, Wada T, et al. Effect of ATRX and G-Quadruplex Formation by the VNTR Sequence on alpha-Globin Gene Expression. Chembiochem. 2016;17(10):928-35.

93. Clynes D, Higgs DR, Gibbons RJ. The chromatin remodeller ATRX: a repeat offender in human disease. Trends Biochem Sci. 2013;38(9):461-6.

94. Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, et al. Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep. 2015;16(7):836-50.

95. Sarma K, Cifuentes-Rojas C, Ergun A, Del Rosario A, Jeon Y, White F, et al. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. Cell. 2014;159(4):869-83.

96. Ren W, Medeiros N, Warneford-Thomson R, Wulfridge P, Yan Q, Bian J, et al. Disruption of ATRX-RNA interactions uncovers roles in ATRX localization and PRC2 function. Nat Commun. 2020;11(1):2219.

97. Eustermann S, Yang JC, Law MJ, Amos R, Chapman LM, Jelinska C, et al. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nat Struct Mol Biol. 2011;18(7):777-82.

98. Ratnakumar K, Duarte LF, LeRoy G, Hasson D, Smeets D, Vardabasso C, et al. ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes Dev. 2012;26(5):433-8.

99. Ratnakumar K, Bernstein E. ATRX: the case of a peculiar chromatin remodeler. Epigenetics. 2013;8(1):3-9.

100. Teng YC, Sundaresan A, O'Hara R, Gant VU, Li M, Martire S, et al. ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. Nat Commun. 2021;12(1):3887.

101. Kernohan KD, Jiang Y, Tremblay DC, Bonvissuto AC, Eubanks JH, Mann MR, et al. ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Dev Cell. 2010;18(2):191-202.

102. Berube NG, Healy J, Medina CF, Wu S, Hodgson T, Jagla M, et al. Patient mutations alter ATRX targeting to PML nuclear bodies. Eur J Hum Genet. 2008;16(2):192-201.

103. Della Ragione F, Vacca M, Fioriniello S, Pepe G, D'Esposito M. MECP2, a multi-talented modulator of chromatin architecture. Brief Funct Genomics. 2016;15(6):420-31.

104. Liyanage VR, Rastegar M. Rett syndrome and MeCP2. Neuromolecular Med. 2014;16(2):231-64.

105. Marano D, Fioriniello S, Fiorillo F, Gibbons RJ, D'Esposito M, Della Ragione F. ATRX Contributes to MeCP2-Mediated Pericentric Heterochromatin Organization during Neural Differentiation. Int J Mol Sci. 2019;20(21):5371.

106. Everett RD, Chelbi-Alix MK. PML and PML nuclear bodies: implications in antiviral defence. Biochimie. 2007;89(6-7):819-30.

107. Tsai K, Thikmyanova N, Wojcechowskyj JA, Delecluse HJ, Lieberman PM. EBV tegument protein BNRF1 disrupts DAXX-ATRX to activate viral early gene transcription. PLoS Pathog. 2011;7(11):e1002376.

108. Delbarre E, Ivanauskiene K, Spirkoski J, Shah A, Vekterud K, Moskaug JO, et al. PML protein organizes heterochromatin domains where it regulates histone H3.3 deposition by ATRX/DAXX. Genome Res. 2017;27(6):913-21.

109. Mahmud I, Liao D. DAXX in cancer: phenomena, processes, mechanisms and regulation. Nucleic Acids Res. 2019;47(15):7734-52.

110. Salomoni P. The PML-interacting protein DAXX: histone loading gets into the picture. Front Oncol. 2013;3:152.

111. Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3. 3. Genes Dev. 2010;24(12):1253-65.

112. Szenker E, Ray-Gallet D, Almouzni G. The double face of the histone variant H3.3. Cell Res. 2011;21(3):421-34.

113. Bano D, Piazzesi A, Salomoni P, Nicotera P. The histone variant H3.3 claims its place in the crowded scene of epigenetics. Aging (Albany NY). 2017;9(3):602-14.

114. Hake SB, Garcia BA, Kauer M, Baker SP, Shabanowitz J, Hunt DF, et al. Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. Proc Natl Acad Sci U S A. 2005;102(18):6344-9.

115. Martire S, Gogate AA, Whitmill A, Tafessu A, Nguyen J, Teng YC, et al. Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. Nat Genet. 2019;51(6):941-6.

116. Frey A, Listovsky T, Guilbaud G, Sarkies P, Sale JE. Histone H3.3 is required to maintain replication fork progression after UV damage. Current biology : CB. 2014;24(18):2195-201.

117. Hinchcliffe EH, Day CA, Karanjeet KB, Fadness S, Langfald A, Vaughan KT, et al. Chromosome missegregation during anaphase triggers p53 cell cycle arrest through histone H3.3 Ser31 phosphorylation. Nat Cell Biol. 2016;18(6):668-75.

118. Amorim JP, Santos G, Vinagre J, Soares P. The Role of ATRX in the Alternative Lengthening of Telomeres (ALT) Phenotype. Genes (Basel). 2016;7(9):66.

119. Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, et al. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS Genet. 2012;8(7):e1002772.

120. Li F, Deng Z, Zhang L, Wu C, Jin Y, Hwang I, et al. ATRX loss induces telomere dysfunction and necessitates induction of alternative lengthening of telomeres during human cell immortalization. EMBO J. 2019;38(19):e96659.

121. Kim J, Sun C, Tran AD, Chin PJ, Ruiz PD, Wang K, et al. The macroH2A1.2 histone variant links ATRX loss to alternative telomere lengthening. Nat Struct Mol Biol. 2019;26(3):213-9.

122. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482(7384):226-31.

123. Chan KM, Fang D, Gan H, Hashizume R, Yu C, Schroeder M, et al. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev. 2013;27(9):985-90.

124. Lewis PW, Müller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science. 2013;340(6134):857-61.

125. Larson JD, Kasper LH, Paugh BS, Jin H, Wu G, Kwon CH, et al. Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression. Cancer Cell. 2019;35(1):140-55 e7.

126. Fang D, Gan H, Lee JH, Han J, Wang Z, Riester SM, et al. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science. 2016;352(6291):1344-8.

127. Jain SU, Khazaei S, Marchione DM, Lundgren SM, Wang X, Weinberg DN, et al. Histone H3.3 G34 mutations promote aberrant PRC2 activity and drive tumor progression. Proc Natl Acad Sci U S A. 2020;117(44):27354-64.

128. Hu J, Hwang SS, Liesa M, Gan B, Sahin E, Jaskelioff M, et al. Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. Cell. 2012;148(4):651-63.

129. Queisser A, Heeg S, Thaler M, von Werder A, Opitz OG. Inhibition of telomerase induces alternative lengthening of telomeres during human esophageal carcinogenesis. Cancer Genet. 2013;206(11):374-86.

130. Active-motif. [Available from: https://www.activemotif.com/documents/Array_Analyze_Software_v16.1.zip.

131. Orchard P, Kyono Y, Hensley J, Kitzman JO, Parker SCJ. Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. Cell Syst. 2020;10(3):298-306 e4.

132. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27(12):1691-2.

133. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

134. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

135. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

136. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014;42(Web Server issue):W187-91.

137. Andrews S. [Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

138. Gordon A, Hannon G. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab cshl edu/fastx_toolkit. 2010;5.
139. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.
140. Becton DaC. [Available from: https://www.flowjo.com/solutions/flowjo/downloads.
141. GraphPad. [Available from: https://www.graphpad.com/scientific-software/prism/.
142. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89.
143. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
144. Bio-rad. [Available from: https://www.bio-rad.com/de-de/product/image-lab-software?ID=KRE6P5E8Z.
145. Arnold K, Gosling J, Holmes D. The Java programming language: Addison Wesley Professional; 2005.
146. Gaspar JM. Improved peak-calling with MACS2. BioRxiv. 2018:496521.
147. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37-52.
148. Desvignes T, Loher P, Eilbeck K, Ma J, Urgese G, Fromm B, et al. Unification of miRNA and isomiR research: the mirGFF3 format and the mirtop API. Bioinformatics. 2020;36(3):698-703.
149. Kang W, Eldfjell Y, Fromm B, Estivill X, Biryukova I, Friedlander MR. miRTrace reveals the organismal origins of microRNA sequencing data. Genome Biol. 2018;19(1):213.
150. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
151. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38(3):276-8.
152. Harshil Patel PE, Alexander Peltzer, Drew Behrens, Gisela Gabernet,Mingda Jin, mashehu, Maxime Garcia. [Available from: https://nf-co.re/atacseq/1.2.1.
153. Steiger S. [Available from: https://github.com/Simontuk/smrnaseq/tree/v1-dev-alt.
154. Institute B. [Available from: https://github.com/broadinstitute/picard.
155. Heger A. [Available from: https://github.com/pysam-developers/pysam.
156. Van Rossum G, Drake Jr FL. Python tutorial: Centrum voor Wiskunde en Informatica Amsterdam; 1995.
157. Team RC. R: A Language and Environment for Statistical Computing.
158. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 2011;12(1):1-16.
159. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184-5.
160. Team R. RStudio: Integrated Development for R.
161. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
162. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.

163. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
164. Krueger F. [Available from: https://github.com/FelixKrueger/TrimGalore.
165. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31(14):2382-3.
166. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
167. Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with 667 enhanced colouring and labeling. R package version 1.6. 0. 668 https://github. com/kevinblighe. EnhancedVolcano; 2020.
168. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.
169. Wickham H. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics. 2011;3(2):180-5.
170. Kassambara A, Kassambara MA. Package 'ggpubr'. 2020.
171. Slowikowski K, Schep A, Hughes S, Lukauskas S, Irisson J-O, Kamvar ZN, et al. Package ggrepel. 2018.
172. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer; 2005. p. 397-420.
173. Bache SM, Wickham H, Bache MSM. Package 'magrittr'. 2020.
174. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14(6):e8124.
175. Dolgalev I. msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. R package version. 2020;7(1).
176. Carlson M, Falcon S, Pages H, Li N. org. Hs. eg. db: Genome wide annotation for Human. R package version. 2019;3(2).
177. Blighe K, Lewis M, Lun A, Blighe MK. Package 'PCAtools'. 2019.
178. Neuwirth E, Neuwirth ME. Package 'RColorBrewer'. CRAN 2011-06-17 08: 34: 00. Apache License 2.0; 2011.
179. Ushey KA, JJ Tang, Yuan. [Available from: https://rstudio.github.io/reticulate/.
180. Shi W, Shi MW, PGM LI, SequenceMatching R, ChIPSeq G, GeneRegulation G. Package 'Rsubread'. 2013.
181. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25(14):1841-2.
182. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686.
183. Team BC, Maintainer BP. TxDb. Hsapiens. UCSC. hg38. knownGene: Annotation package for TxDb object (s). R package. 2019.
184. Hoang SM, O'Sullivan RJ. Alternative Lengthening of Telomeres: Building Bridges To Connect Chromosome Ends. Trends Cancer. 2020;6(3):247-60.
185. Braun DM, Chung I, Kepper N, Deeg KI, Rippe K. TelNet-a database for human and yeast genes involved in telomere maintenance. BMC Genetics. 2018;19(1):1-10.
186. Henson JD, Hannay JA, McCarthy SW, Royds JA, Yeager TR, Robinson RA, et al. A robust assay for alternative lengthening of telomeres in tumors shows the significance of alternative lengthening of telomeres in sarcomas and astrocytomas. Clin Cancer Res. 2005;11(1):217-25.

187. Shang W-H, Hori T, Westhorpe FG, Godek KM, Toyoda A, Misu S, et al. Acetylation of histone H4 lysine 5 and 12 is required for CENP-A deposition into centromeres. Nat Commun. 2016;7(1):1-13.

188. Roider HG, Manke T, O'Keeffe S, Vingron M, Haas SA. PASTAA: identifying transcription factors associated with sets of co-regulated genes. Bioinformatics. 2009;25(4):435-42.

189. Kern F, Fehlmann T, Solomon J, Schwed L, Grammes N, Backes C, et al. miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. Nucleic Acids Res. 2020;48(W1):W521-W8.

190. Deeg KI, Chung I, Poos AM, Braun DM, Korshunov A, Oswald M, et al. Dissecting telomere maintenance mechanisms in pediatric glioblastoma. BioRxiv. 2017:129106.

191. Mason-Osann E, Dai A, Floro J, Lock YJ, Reiss M, Gali H, et al. Identification of a novel gene fusion in ALT positive osteosarcoma. Oncotarget. 2018;9(67):32868-80.

192. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23(4):183-91.

193. Droin NM, Pinkoski MJ, Dejardin E, Green DR. Egr family members regulate nonlymphoid expression of Fas ligand, TRAIL, and tumor necrosis factor during immune responses. Mol Cell Biol. 2003;23(21):7638-47.

194. Lafferty-Whyte K, Bilsland A, Hoare SF, Burns S, Zaffaroni N, Cairney CJ, et al. TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT expression. Neoplasia. 2010;12(5):405-14.

195. Szczepny A, Carey K, McKenzie L, Jayasekara WSN, Rossello F, Gonzalez-Rajal A, et al. The tumor suppressor Hic1 maintains chromosomal stability independent of Tp53. Oncogene. 2018;37(14):1939-48.

196. Verhoeven Y, Tilborghs S, Jacobs J, De Waele J, Quatannens D, Deben C, et al., editors. The potential and controversy of targeting STAT family members in cancer. Semin Cancer Biol; 2020: Elsevier.

197. Pan MG, Xiong Y, Chen F. NFAT gene family in inflammation and cancer. Curr Mol Med. 2013;13(4):543-54.

198. Cerone MA, Burgess DJ, Naceur-Lombardelli C, Lord CJ, Ashworth A. High-throughput RNAi screening reveals novel regulators of telomerase. Cancer Res. 2011;71(9):3328-40.

199. Giannone RJ, McDonald HW, Hurst GB, Shen RF, Wang Y, Liu Y. The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. PLoS One. 2010;5(8):e12407.

200. Ramlee MK, Wang J, Toh WX, Li S. Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene. Genes (Basel). 2016;7(8):50.

201. Tarca AL, Draghici S, Bhatti G, Romero R. Down-weighting overlapping genes improves gene set analysis. BMC Bioinform. 2012;13(1):136.

202. Miao H-Q, Klagsbrun M. Neuropilin is a mediator of angiogenesis. Cancer Metastasis Rev. 2000;19(1):29-37.

203. Li N, You X, Chen T, Mackowiak SD, Friedländer MR, Weigt M, et al. Global profiling of miRNAs and the hairpin precursors: insights into miRNA processing and novel miRNA discovery. Nucleic Acids Res. 2013;41(6):3619-34.

204. Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. Nat Rev Genet. 2020;21(12):721-36.

205. Mevel R, Draper JE, Lie ALM, Kouskoff V, Lacaud G. RUNX transcription factors: orchestrators of development. Development. 2019;146(17):dev148296.

206. Lefebvre V, Dumitriu B, Penzo-Mendez A, Han Y, Pallavi B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol. 2007;39(12):2195-214.
207. Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. Blood. 2017;129(15):2070-82.
208. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci. 2014;35(2):76-85.
209. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer. 2003;3(11):859-68.
210. Atsaves V, Leventaki V, Rassidakis GZ, Claret FX. AP-1 transcription factors as regulators of immune responses in cancer. Cancers. 2019;11(7):1037.
211. Reczek CR, Chandel NS. The two faces of reactive oxygen species in cancer. Annu Rev Cancer Biol. 2017;1:79-98.
212. Currie E, Schulze A, Zechner R, Walther TC, Farese RV, Jr. Cellular fatty acid metabolism and cancer. Cell Metab. 2013;18(2):153-61.
213. Niizuma K, Endo H, Chan PH. Oxidative stress and mitochondrial dysfunction as determinants of ischemic neuronal death and survival. J Neurochem. 2009;109:133-8.
214. Lin Y, Jiang M, Chen W, Zhao T, Wei Y. Cancer and ER stress: Mutual crosstalk between autophagy, oxidative stress and inflammatory response. Biomed Pharmacother. 2019;118:109249.
215. Flynn RL, Cox KE, Jeitany M, Wakimoto H, Bryll AR, Ganem NJ, et al. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. Science. 2015;347(6219):273-7.
216. Chen Y-A, Shen Y-L, Hsia H-Y, Tiang Y-P, Sung T-L, Chen L-Y. Extrachromosomal telomere repeat DNA is linked to ALT development via cGAS-STING DNA sensing pathway. Nat Struct Mol. 2017;24(12):1124-31.
217. Wiser C, Kim B, Vincent J, Ascano M. Small molecule inhibition of human cGAS reduces total cGAMP output and cytokine expression in cells. Sci Rep. 2020;10(1):7604.
218. Hertzog J, Rehwinkel J. Regulation and inhibition of the DNA sensor cGAS. EMBO Rep. 2020;21(12):e51345.
219. Mohammad F, Weissmann S, Leblanc B, Pandey DP, Højfeldt JW, Comet I, et al. EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. Nat Med. 2017;23(4):483-92.
220. Heaphy CM, Subhawong AP, Hong SM, Goggins MG, Montgomery EA, Gabrielson E, et al. Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. Am J Pathol. 2011;179(4):1608-15.
221. Henson JD, Lau LM, Koch S, Martin La Rotta N, Dagg RA, Reddel RR. The C-Circle Assay for alternative-lengthening-of-telomeres activity. Methods. 2017;114:74-84.
222. Liu H, Xie Y, Zhang Z, Mao P, Liu J, Ma W, et al. Telomeric Recombination Induced by DNA Damage Results in Telomere Extension and Length Heterogeneity. Neoplasia. 2018;20(9):905-16.
223. Raghunandan M, Geelen D, Majerova E, Decottignies A. hTR and NHP2 independently activate DNA damage response at ALT telomeres. The EMBO J. 2021.
224. Doksani Y. The response to DNA damage at telomeric repeats and its consequences for telomere function. Genes. 2019;10(4):318.

225. Schmitt CA, Fridman JS, Yang M, Lee S, Baranov E, Hoffman RM, et al. A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy. Cell. 2002;109(3):335-46.

226. Buczkowicz P, Bartels U, Bouffet E, Becher O, Hawkins C. Histopathological spectrum of paediatric diffuse intrinsic pontine glioma: diagnostic and therapeutic implications. Acta Neuropathol. 2014;128(4):573-81.

227. Muller AM, Florek M. 5-Azacytidine/5-Azacitidine. Recent Results Cancer Res. 2014;201:299-324.

228. Banerjee S, Gusho E, Gaughan C, Dong B, Gu X, Holvey-Bates E, et al. OAS-RNase L innate immune pathway mediates the cytotoxicity of a DNA-demethylating drug. Proc Natl Acad Sci U S A. 2019;116(11):5071-6.

229. Ng LJ, Cropley JE, Pickett HA, Reddel RR, Suter CM. Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. Nucleic Acids Res. 2009;37(4):1152-9.

230. Kim W, Ludlow AT, Min J, Robin JD, Stadler G, Mender I, et al. Regulation of the Human Telomerase Gene TERT by Telomere Position Effect-Over Long Distances (TPE-OLD): Implications for Aging and Cancer. PLoS Biol. 2016;14(12):e2000016.

231. Li M, Dong Q, Zhu B. Aurora kinase B phosphorylates histone H3. 3 at serine 31 during mitosis in mammalian cells. J Mol Bio. 2017;429(13):2042-5.

232. Ricke RM, Jeganathan KB, Malureanu L, Harrison AM, Van Deursen JM. Bub1 kinase activity drives error correction and mitotic checkpoint control but not tumor suppression. J Cell Biol. 2012;199(6):931-49.

233. Zhang QS, Manche L, Xu RM, Krainer AR. hnRNP A1 associates with telomere ends and stimulates telomerase activity. RNA. 2006;12(6):1116-28.

234. Dai J, Sultan S, Taylor SS, Higgins JM. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. Genes Dev. 2005;19(4):472-88.

235. Wang F, Ulyanova NP, van der Waal MS, Patnaik D, Lens SM, Higgins JM. A positive feedback loop involving Haspin and Aurora B promotes CPC accumulation at centromeres in mitosis. Current biology : CB. 2011;21(12):1061-9.

236. Lawlor RT, Veronese N, Pea A, Nottegar A, Smith L, Pilati C, et al. Alternative lengthening of telomeres (ALT) influences survival in soft tissue sarcomas: a systematic review with meta-analysis. BMC Cancer. 2019;19(1):232.

237. Sung J-Y, Lim H-W, Joung J-G, Park W-Y. Pan-Cancer Analysis of Alternative Lengthening of Telomere Activity. Cancers. 2020;12(8):2207.

238. Chang K, Creighton C, Davis C, Donehower L. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45(10):1113-20.

239. Pang L, Li B, Zheng B, Niu L, Ge L. miR-138 inhibits gastric cancer growth by suppressing SOX4. Oncol Rep. 2017;38(2):1295-302.

240. Zhang J, Zhang K, Hou Y. Long noncoding RNA NNTAS1 knockdown represses the progression of gastric cancer via modulating the miR1425p/SOX4/Wnt/betacatenin signaling pathway. Mol Med Rep. 2020;22(2):687-96.

241. Zhang Y, Xie RL, Croce CM, Stein JL, Lian JB, van Wijnen AJ, et al. A program of microRNAs controls osteogenic lineage progression by targeting transcription factor Runx2. Proc Natl Acad Sci U S A. 2011;108(24):9863-8.

242. D'Alonzo RC, Selvamurugan N, Karsenty G, Partridge NC. Physical interaction of the activator protein-1 factors c-Fos and c-Jun with Cbfa1 for collagenase-3 promoter activation. J Biol Chem. 2002;277(1):816-22.
243. Hess J, Porte D, Munz C, Angel P. AP-1 and Cbfa/runt physically interact and regulate parathyroid hormone-dependent MMP13 expression in osteoblasts through a new osteoblast-specific element 2/AP-1 composite element. J Biol Chem. 2001;276(23):20029-38.
244. Korinfskaya S, Parameswaran S, Weirauch MT, Barski A. Runx Transcription Factors in T Cells—What Is Beyond Thymic Development? Front Immunol. 2021;12.
245. Kyriakis JM. Activation of the AP-1 transcription factor by inflammatory cytokines of the TNF family. Gene Expr. 1999;7(4-6):217-31.
246. O'Sullivan RJ, Arnoult N, Lackner DH, Oganesian L, Haggblom C, Corpet A, et al. Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1. Nat Struct Mol Biol. 2014;21(2):167-74.
247. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436-44.
248. Cron RQ, Bort SJ, Wang Y, Brunvand MW, Lewis DB. T cell priming enhances IL-4 gene expression by increasing nuclear factor of activated T cells. J Immunol. 1999;162(2):860-70.
249. Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB. Oxidative stress, inflammation, and cancer: how are they linked? Free Radic Biol Med. 2010;49(11):1603-16.
250. von Zglinicki T. Oxidative stress shortens telomeres. Trends Biochem Sci. 2002;27(7):339-44.
251. Reichmann D, Voth W, Jakob U. Maintaining a Healthy Proteome during Oxidative Stress. Mol Cell. 2018;69(2):203-13.
252. Wang Y, Wan B, Li D, Zhou J, Li R, Bai M, et al. BRSK2 is regulated by ER stress in protein level and involved in ER stress-induced apoptosis. Biochem Biophys Res Commun. 2012;423(4):813-8.
253. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis. 2009;30(7):1073-81.
254. Grivennikov SI, Karin M. Inflammation and oncogenesis: a vicious connection. Curr Opin Genet Dev. 2010;20(1):65-71.
255. Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol. 2017;14(12):717-34.
256. Sharma P, Allison JP. The future of immune checkpoint therapy. Science. 2015;348(6230):56-61.
257. Lin S-Y, Morrison JR, Phillips DJ, de Kretser DM. Regulation of ovarian function by the TGF-b superfamily and follistatin. Reproduction. 2003;126(2):133-48.
258. Chiarle R, Simmons WJ, Cai H, Dhall G, Zamo A, Raz R, et al. Stat3 is required for ALK-mediated lymphomagenesis and provides a possible therapeutic target. Nat Med. 2005;11(6):623-9.
259. Seystahl K, Tritschler I, Szabo E, Tabatabai G, Weller M. Differential regulation of TGF-beta-induced, ALK-5-mediated VEGF release by SMAD2/3 versus SMAD1/5/8 signaling in glioblastoma. Neuro Oncol. 2015;17(2):254-65.
260. Chen YG, Lui HM, Lin SL, Lee JM, Ying SY. Regulation of cell proliferation, apoptosis, and carcinogenesis by activin. Exp Biol Med (Maywood). 2002;227(2):75-87.
261. Hedger MP, Winnall WR, Phillips DJ, de Kretser DM. The regulation and functions of activin and follistatin in inflammation and immunity. Vitam Horm. 2011;85:255-97.

262. Phillips DJ, de Kretser DM, Hedger MP. Activin and related proteins in inflammation: not just interested bystanders. Cytokine Growth Factor Rev. 2009;20(2):153-64.

263. Robertson DM, Burger HG, Fuller PJ. Inhibin/activin and ovarian cancer. Endocr Relat Cancer. 2004;11(1):35-49.

264. de Winter JP, ten Dijke P, de Vries CJ, van Achterberg TA, Sugino H, de Waele P, et al. Follistatins neutralize activin bioactivity by inhibition of activin binding to its type II receptors. Mol Cell Endocrinol. 1996;116(1):105-14.

265. Sakurai A, Doci CL, Gutkind JS. Semaphorin signaling in angiogenesis, lymphangiogenesis and cancer. Cell Res. 2012;22(1):23-32.

266. Ellis LM. The role of neuropilins in cancer. Mol Cancer Ther. 2006;5(5):1099-107.

267. Sia D, Alsinet C, Newell P, Villanueva A. VEGF signaling in cancer treatment. Curr Pharm Des. 2014;20(17):2834-42.

268. Fang W, Hartmann N, Chow D, Riegel A, Wellstein A. Pleiotrophin stimulates fibroblasts and endothelial and epithelial cells and is expressed in human cancer. J Biol Chem. 1992;267(36):25889-97.

269. Kong Y, Bai P-s, Nan K-j, Sun H, Chen N-z, Qi X-g. Pleiotrophin is a potential colorectal cancer prognostic factor that promotes VEGF expression and induces angiogenesis in colorectal cancer. Int J Colorectal Dis. 2012;27(3):287-98.

270. Wang HQ, Wang J. Expression of pleiotrophin in small cell lung cancer. J Biol Regul Homeost Agents. 2015;29(1):175-9.

271. Lynn KD, Roland CL, Brekken RA. VEGF and pleiotrophin modulate the immune profile of breast cancer. Cancers (Basel). 2010;2(2):970-88.

272. Fukada M, Fujikawa A, Chow JP, Ikematsu S, Sakuma S, Noda M. Protein tyrosine phosphatase receptor type Z is inactivated by ligand-induced oligomerization. FEBS Letters. 2006;580(17):4051-6.

273. Mikelis C, Sfaelou E, Koutsioumpa M, Kieffer N, Papadimitriou E. Integrin αvβ3 is a pleiotrophin receptor required for pleiotrophin-induced endothelial cell migration through receptor protein tyrosine phosphatase β/ζ. FASEB J. 2009;23(5):1459-69.

274. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in cancer. Vasc Health Risk Manag. 2006;2(3):213-9.

275. Manda G, Nechifor MT, Neagu T-M. Reactive oxygen species, cancer and anti-cancer therapies. Curr Chem BIol. 2009;3(1):22-46.

276. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne). 2018;9:402.

277. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148-D54.

278. Arif K, Elliott EK, Haupt LM, Griffiths LR. Regulatory mechanisms of epigenetic miRNA relationships in human cancer and potential as therapeutic targets. Cancers. 2020;12(10):2922.

279. Kaul Z, Cheung CTY, Bhargava P, Sari AN, Yu Y, Huifu H, et al. Functional characterization of miR-708 microRNA in telomerase positive and negative human cancer cells. Sci Rep. 2021;11(1):17052.

280. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol. 2015;7(4):a016600.

281. Li J, Lei H, Xu Y, Tao ZZ. miR-512-5p suppresses tumor growth by targeting hTERT in telomerase positive head and neck squamous cell carcinoma in vitro and in vivo. PLoS One. 2015;10(8):e0135265.

282. Hsu P-S, Yu S-H, Tsai Y-T, Chang J-Y, Tsai L-K, Ye C-H, et al. More than causing (epi) genomic instability: emerging physiological implications of transposable element modulation. J Biol Chem. 2021;28(1):1-14.

283. Leibold DM, Swergold GD, Singer MF, Thayer RE, Dombroski BA, Fanning TG. Translation of LINE-1 DNA elements in vitro and in human cells. Proc Natl Acad Sci U S A. 1990;87(18):6990-4.

284. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 2000;24(4):363-7.

285. Rodić N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol. 2014;184(5):1280-6.

286. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019;20(2):89-108.

287. Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell. 2012;149(3):693-707.

288. Fu A, Jacobs DI, Hoffman AE, Zheng T, Zhu Y. PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. Carcinogenesis. 2015;36(10):1094-102.

289. Weng W, Liu N, Toiyama Y, Kusunoki M, Nagasaka T, Fujiwara T, et al. Novel evidence for a PIWI-interacting RNA (piRNA) as an oncogenic mediator of disease progression, and a potential prognostic biomarker in colorectal cancer. Mol Cancer. 2018;17(1):1-12.

290. Pek JW, Anand A, Kai T. Tudor domain proteins in development. Development. 2012;139(13):2255-66.

291. Teissandier A, Servant N, Barillot E, Bourc’his D. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. Mobile DNA. 2019;10(1):1-12.

292. Cheng Y, Ma Z, Kim B-H, Wu W, Cayting P, Boyle AP, et al. Principles of regulatory information conservation between mouse and human. Nature. 2014;515(7527):371-5.

Acknowledgments

I would like to thank the many people that have, in their own special way, supported me on this PhD journey.

First of all, thank you Karsten for accepting me into your lab, and allowing me to explore new methods that I had never thought I could do. Thank you for your trust and your ever-present can-do attitude that has motivated me throughout this thesis.

Thank you to my TAC members, Sevin Turcan and Frank Westermann for your continued support during this thesis, the many helpful discussions, and for being part of my defense committee. I would additionally like to thank Rob Russell for chairing the defense committee.

Thank you to the core facilities in DKFZ, namely the flow cytometry, genomics and light microscopy, for your support.

I would like to thank my collaboration partners, Eran Meshorer, whom we worked with as part of the DKFZ-MOST project, and Priya Chudasama for the C-circle screening project.

I would like to thank all past and present members of the Rippe lab. Afzal, Ahmad, Alexandra, Anne, Caro, Delia, Emma, Fabian, Isabelle, Jana, Jorge, Kathi, Lara, Linda, Lukas, Michelle, Nick, Norbert, Philipp, Robin, Sabrina, Simon, Sofie, Stephan, Verena. Whether it was just complaining about failed experiments while waiting for coffee or helping me with experiments and analysis, I'm very grateful for your support! I will always remember all the after-work activities fondly, like playing cards against humanity and what do you meme or making the craziest shapes from glowsticks when we were in Urban Kitchen. The atmosphere in the lab was always so encouraging (and light-hearted and fun), and I'm happy to have worked with you. I would like to thank the "Telo-Tubbies" particularly: Lukas, Caro, Emma and formerly Alexandra and Delia, for your many helpful discussions about ALT, and what is it good for? Thanks to Caro for helping me with the sequencing and metaphase spreads, and thanks Sabrina for your help with setting up the ATAC-seq. For helping me click the enter button powerfully enough, thank you Markus, Lara, Philipp, Isabelle, Simon, Linda, Lukas, Jorge. I now realized that I should also press tab sometimes! Special thanks to Markus for helping me with DEseq, Lara for ATAC-related matters, Isabelle who helped me with the MOFA analysis, and Simon for the many IT-related issues and setting up the NF-core pipelines. Additional thanks to Lukas and Markus who proofread my thesis and gave me many helpful comments. Thanks to Sofie for your help with all administrative matters.

I would also want to thank the labs of Kristian Dreij, Boris Zhivotovsky and Richard Wade-Martins, where I came into contact with many different areas of science. I would particularly like to thank my Master thesis supervisor, Birce Akpinar, a kick-a** scientist who inspired me to continue on with a doctorate.

Jelena (aka Puž aka flera andra namn haha!). We've been friends for more than 20 years, since elementary school!! Thank you for making me feel very old! And also thank you for always having the time to see me whenever I stop by Malmö, to chat, drink good coffee and answering :"When will you cure cancer?".

Laila, we have also been friends for 20 years! We always had a laugh about silly things and thank you for always being able to talk/complain about everything possible. Thank you for your friendship and the great food (and Kexchoklad) that I always looked forward to when coming home.

Lotta & Alex. We moved up to Stockholm when we were 18 (I'm not sure how we survived!). Thank you for that great time! From the Tallinn-cruise ages ago to celebrating Valborg in Uppsala, some great (very hazy!) memories were made.

To Sam, Siobhan and Jean. As your Swedish "buddy" I think I failed to show you any Swedish things. However, it was great to meet you during my Bachelor's and we had lots of fun! I'm looking forward to going back to the Video Game Museum one day with you all.

To Naz, the chilliest people I know. Thank you for being a great friend to me, and also for the "Mörby Nights". Watching Doctor Who, being healthy and cooking fish, really helped me during the stressful times of writing a Master thesis.

To the people that I met during my Master's, thank you Anna, Therese, Lucia, Sara, Sarah, Manon. Although we are scattered a bit everywhere in Europe now, I feel like I have friends in every city!

To the people that I met in Oxford, Athena, Angela, Johannes, Tara, Juli, Sarah, Felix, Lea and Ron. Going on Formals and living the Oxford College life, having brunch, or being roommates was a fun experience, thank you!

And for the people that made my time in Heidelberg great. Thank you, Anita, Agnes, Ahmad, Joschka and Linda. Watching fail videos, going to the Neckar, or just chilling in the garden was really life-saving during the stressful PhD times.

To my family, Branimirka, Adnan, Nina, Jonas, Joar, Laila and Alva. You supported me throughout, and taught me to be stubborn, kind, and never giving up! Although I've been far away, and especially now during the pandemic, I'm always happy to go back home.

And last, but not least. Thank you, Jonny, for being in my life. You have always supported me, and when I thought I couldn't do things you always told me the opposite. From proof-reading my thesis, to discuss lab results late at night, to taking care of housework when I was writing, and sending fun memes, you have always been there for me. Without you I don't think I could have done what I have, so thank you!

Appendix

I. Cell lines and sequencing IDs

Supplemental Table 1. ILSE IDs for each sequencing experiment.

Tumor entity	Cell line	Replicate	Long RNA-seq ID	Small RNA-seq ID	ATAC-seq ID	
					R1	R2
Pediatric glioblastoma	KNS42	1	AS-512665-LR-52219_R1	AS-512637-LR-52459_R1	AS-535533-LR-53335_R1	AS-535533-LR-53335_R2
					AS-535510-LR-53334_R1	AS-535510-LR-53334_R2
		2	AS-512706-LR-52335_R1	AS-565990-LR-55484_R1	AS-535510-LR-53335_R1	AS-535510-LR-53335_R2
					AS-535535-LR-53334_R1	AS-535535-LR-53334_R2
	MGMB1	1	AS-512667-LR-52219_R1	AS-512639-LR-52459_R1	AS-535519-LR-53334_R1	AS-535519-LR-53334_R2
					AS-535519-LR-53335_R1	AS-535519-LR-53335_R2
		2	AS-512708-LR-52335_R1	AS-512729-LR-52870_R1	AS-535543-LR-53334_R1	AS-535543-LR-53334_R2
					AS-535543-LR-53335_R1	AS-535543-LR-53335_R2
	SF188	1	AS-512662-LR-52218_R1	AS-512634-LR-52458_R1	AS-535521-LR-53334_R1	AS-535521-LR-53334_R2
					AS-535521-LR-53335_R1	AS-535521-LR-53335_R2
		2	AS-512703-LR-52334_R1	AS-512724-LR-52869_R1	AS-535545-LR-53334_R1	AS-535545-LR-53334_R2
					AS-535545-LR-53335_R1	AS-535545-LR-53335_R2
	NEM157	1	AS-512669-LR-52219_R1	AS-512641-LR-52459_R1	AS-535523-LR-53334_R1	AS-535523-LR-53334_R2
					AS-535523-LR-53335_R1	AS-535523-LR-53335_R2
		2	AS-512710-LR-52335_R1	AS-512731-LR-52870_R1	AS-535547-LR-53334_R1	AS-535547-LR-53334_R2
					AS-535547-LR-53335_R1	AS-535547-LR-53335_R2
	NEM165	1	AS-512674-LR-52331_R1	AS-565994-LR-55483_R1	AS-535525-LR-53334_R1	AS-535525-LR-53334_R2
					AS-535525-LR-53335_R1	AS-535525-LR-53335_R2
		2	AS-512683-LR-52332_R1	AS-565992-LR-55483_R1	AS-535549-LR-53334_R1	AS-535549-LR-53334_R2
					AS-535549-LR-53335_R1	AS-535549-LR-53335_R2
	NEM168	1	AS-512676-LR-52331_R1	AS-512646-LR-52460_R1	AS-535527-LR-53334_R1	AS-535527-LR-53334_R2
					AS-535527-LR-53335_R1	AS-535527-LR-53335_R2
		2	AS-512685-LR-52332_R1	AS-512736-LR-52871_R1	AS-535551-LR-53334_R1	AS-535551-LR-53334_R2
					AS-535551-LR-53335_R1	AS-535551-LR-53335_R2
	NEM168 ATRX -/- clone B5	1	AS-512678-LR-52331_R1	AS-512648-LR-52460_R1	AS-535529-LR-53334_R1	AS-535529-LR-53334_R2
					AS-535529-LR-53335_R1	AS-535529-LR-53335_R2
		2	AS-512687-LR-52332_R1	AS-512738-LR-52871_R1	AS-535553-LR-53334_R1	AS-535553-LR-53334_R2
					AS-535553-LR-53335_R1	AS-535553-LR-53335_R2
Osteosarcoma	HOS	1	AS-512653-LR-52217_R1	AS-512625-LR-52457_R1	AS-535508-LR-53334_R1	AS-535508-LR-53334_R2
					AS-535508-LR-53335_R1	AS-535508-LR-53335_R2
		2	AS-512694-LR-52333_R1	AS-512715-LR-52868_R1	AS-535533-LR-53334_R1	AS-535533-LR-53334_R2
					AS-535533-LR-53335_R1	AS-535533-LR-53335_R2
	MG.63	1	AS-512655-LR-52217_R1	AS-512627-LR-52457_R1	AS-535516-LR-53334_R1	AS-535516-LR-53334_R2
					AS-535516-LR-53335_R1	AS-535516-LR-53335_R2
		2	AS-512696-LR-52333_R1	AS-512717-LR-52868_R1	AS-535541-LR-53334_R1	AS-535541-LR-53334_R2
					AS-535541-LR-53335_R1	AS-535541-LR-53335_R2
	Cal72	1	AS-512660-LR-52218_R1	AS-512632-LR-52458_R1	AS-535512-LR-53334_R1	AS-535512-LR-53334_R2
					AS-535512-LR-53335_R1	AS-535512-LR-53335_R2
		2	AS-512701-LR-52334_R1	AS-512722-LR-52869_R1	AS-535537-LR-53334_R1	AS-535537-LR-53334_R2
					AS-535537-LR-53335_R1	AS-535537-LR-53335_R2
	Saos-2	1	AS-512658-LR-52218_R1	AS-512630-LR-52458_R1	AS-535514-LR-53334_R1	AS-535514-LR-53334_R2
					AS-535514-LR-53335_R1	AS-535514-LR-53335_R2
		2	AS-512689-LR-52332_R1, AS-512699-LR-52334_R1, AS-512699-LR-52334_R1	AS-565988-LR-55484_R1	AS-535539-LR-53334_R1	AS-535539-LR-53334_R2
					AS-535539-LR-53335_R1	AS-535539-LR-53335_R2
	U2OS	1	AS-512651-LR-52217_R1, AS-512680-LR-52331_R1	AS-512623-LR-52457_R1	AS-535506-LR-53334_R1	AS-535506-LR-53334_R2
					AS-535506-LR-53335_R1	AS-535506-LR-53335_R2
		2	AS-512692-LR-52333_R1	AS-565986-LR-55483_R1	AS-535531-LR-53334_R1	AS-535531-LR-53334_R2
					AS-535531-LR-53335_R1	AS-535531-LR-53335_R2

II. Differentially expressed genes adjusted p-value > $1 \cdot 10^{10}$

Supplemental Table 2. All differentially expressed genes when comparing ALT+ and ALT- cell lines with an adjusted p-value above $1 \cdot 10^{10}$.

Ensembl ID	Gene name	baseMean	log2FoldChange	pvalue	padj
ENSG00000105894	PTN	882.9065	7.788914	2.64E-60	9.97E-56
ENSG00000280709	LINC02203	49.13565	24.55739	1.20E-40	2.26E-36
ENSG00000121388	AL139327.1	53.4065	24.69942	2.30E-38	2.90E-34
ENSG00000122641	INHBA	7705.56	7.205526	9.14E-36	8.64E-32
ENSG00000174672	BRSK2	293.1548	-2.77144	6.71E-26	5.07E-22
ENSG00000184058	TBX1	2036.707	-7.66955	6.11E-25	3.85E-21
ENSG00000199798	SNORD114-5	209.2706	26.22195	3.36E-23	1.82E-19
ENSG00000283217	AC068205.1	35.03619	23.4611	5.63E-23	2.66E-19
ENSG00000051523	CYBA	1625.359	-6.69797	5.45E-22	2.29E-18
ENSG00000149256	TENM4	6145.347	7.488519	2.04E-21	7.30E-18
ENSG00000176658	MYO1D	2474.749	5.838109	2.13E-21	7.30E-18
ENSG00000116183	PAPPA2	895.9513	7.529534	6.06E-21	1.91E-17
ENSG00000118785	SPP1	4061.356	9.97863	3.56E-20	1.03E-16
ENSG00000118946	PCDH17	406.5143	6.698643	7.09E-20	1.91E-16
ENSG00000060140	STYK1	97.14748	5.245483	7.82E-20	1.97E-16
ENSG00000129596	CDO1	141.3338	5.204402	1.58E-18	3.73E-15
ENSG00000154678	PDE1C	4685.581	6.23908	3.07E-18	6.52E-15
ENSG00000140538	NTRK3	1623.736	8.225666	3.11E-18	6.52E-15
ENSG00000132329	RAMP1	940.3237	-4.29955	1.55E-17	3.08E-14
ENSG00000178033	CALHM5	416.6219	3.376114	1.94E-17	3.67E-14
ENSG00000011677	GABRA3	873.3832	7.345791	2.24E-17	4.02E-14
ENSG00000157680	DGKI	554.2693	4.656007	4.43E-17	7.47E-14
ENSG00000103710	RASL12	748.1367	-7.70457	4.55E-17	7.47E-14
ENSG00000169515	CCDC8	662.1255	-7.85399	6.18E-17	9.73E-14
ENSG00000149292	TTC12	344.2773	-9.5901	7.68E-17	1.16E-13
ENSG00000180543	TSPYL5	1865.298	9.363672	9.05E-17	1.32E-13
ENSG00000188620	HMX3	432.0144	-7.48066	9.92E-17	1.39E-13
ENSG00000188848	BEND4	1657.857	-6.91559	1.52E-16	2.05E-13
ENSG00000196581	AJAP1	6723.008	9.029328	1.79E-16	2.34E-13
ENSG00000130675	MNX1	180.9574	-7.34707	2.63E-16	3.31E-13
ENSG00000113494	PRLR	121.3976	4.453885	2.81E-16	3.43E-13
ENSG00000176293	ZNF135	190.9945	5.372333	3.37E-16	3.98E-13
ENSG00000197249	SERPINA1	582.0168	7.165605	4.10E-16	4.70E-13
ENSG00000154646	TMPRSS15	818.6865	7.631573	7.79E-16	8.66E-13
ENSG00000150051	MKX	2123.662	-5.42205	9.10E-16	9.82E-13
ENSG00000156968	MPV17L	660.1479	-2.32633	1.32E-15	1.39E-12
ENSG00000221866	PLXNA4	4114.911	6.498053	1.89E-15	1.93E-12
ENSG00000280241	AC079298.3	20.25047	7.590574	2.20E-15	2.19E-12
ENSG00000102755	FLT1	376.919	-4.9778	2.59E-15	2.51E-12
ENSG00000165071	TMEM71	136.0187	6.050602	3.08E-15	2.87E-12
ENSG00000110665	C11orf21	83.33669	-5.11051	3.11E-15	2.87E-12
ENSG00000153707	PTPRD	781.8681	6.876697	4.76E-15	4.28E-12
ENSG00000069482	GAL	290.4932	-7.40348	5.87E-15	5.16E-12
ENSG00000154556	SORBS2	491.0218	4.65777	8.38E-15	7.03E-12
ENSG00000268388	FENDRR	4196.511	-8.17881	8.29E-15	7.03E-12
ENSG00000196154	S100A4	5116.903	-3.72743	1.14E-14	9.36E-12

ENSG00000236969	GGT8P	96.03426	3.950628	1.26E-14	1.01E-11
ENSG00000141052	MYOCD	525.4517	7.138946	1.81E-14	1.43E-11
ENSG00000172260	NEGR1	1910.198	4.153833	1.87E-14	1.44E-11
ENSG00000181072	CHRM2	55.83296	8.326651	1.98E-14	1.50E-11
ENSG00000229807	XIST	6390.972	8.566679	2.87E-14	2.13E-11
ENSG00000135549	PKIB	535.7891	6.708132	3.66E-14	2.66E-11
ENSG00000138356	AOX1	2843.203	4.260794	5.50E-14	3.92E-11
ENSG00000230500	MKX-AS1	14.85446	-5.95641	7.68E-14	5.37E-11
ENSG00000143013	LMO4	3780.748	3.1356	9.87E-14	6.74E-11
ENSG00000224732	MAGEA7P	31.25368	6.264835	9.99E-14	6.74E-11
ENSG00000158125	XDH	218.8856	4.985375	1.04E-13	6.91E-11
ENSG00000196460	RFX8	350.0143	6.307928	1.18E-13	7.68E-11
ENSG00000170954	ZNF415	241.9737	4.008162	1.38E-13	8.85E-11
ENSG00000152208	GRID2	245.1863	8.032148	1.47E-13	9.27E-11
ENSG00000206557	TRIM71	129.6151	-6.33977	1.66E-13	1.03E-10
ENSG00000183671	GPR1	615.7035	5.030572	1.71E-13	1.04E-10
ENSG00000221867	MAGEA3	2069.183	7.349727	2.10E-13	1.26E-10
ENSG00000164161	HHIP	1080.792	5.030536	2.93E-13	1.73E-10
ENSG00000154319	FAM167A	1547.834	4.199812	3.16E-13	1.84E-10
ENSG00000176049	JAKMIP2	613.9904	4.969614	3.67E-13	2.09E-10
ENSG00000164687	FABP5	1229.525	-5.40989	3.71E-13	2.09E-10
ENSG00000159674	SPON2	847.5757	-3.54659	4.38E-13	2.43E-10
ENSG00000203930	LINC00632	5863.406	4.884529	5.02E-13	2.75E-10
ENSG00000231114	AC078842.2	10.00695	6.589997	5.81E-13	3.14E-10
ENSG00000235385	LINC02154	148.1644	5.711891	6.37E-13	3.39E-10
ENSG00000180347	CCDC129	43.56958	6.842791	7.70E-13	3.99E-10
ENSG00000069431	ABCC9	1502.059	6.62949	7.70E-13	3.99E-10
ENSG00000189001	SBSN	360.5687	6.563917	8.43E-13	4.31E-10
ENSG00000021355	SERPINB1	852.3861	-4.72981	9.80E-13	4.94E-10
ENSG00000118473	SGIP1	615.8778	7.177536	1.51E-12	7.49E-10
ENSG00000164778	EN2	309.0029	-7.42978	1.64E-12	8.03E-10
ENSG00000160886	LY6K	592.2258	-4.20583	1.91E-12	9.25E-10

baseMean , average normalized count values which have in turn been normalized to library size factor; Log2FoldChange, average fold change expressed in Log2; p-value, non-adjusted average p-value for gene; padj, average Benjamini-Hochberg adjusted p-value.

III. miRNA differential expression

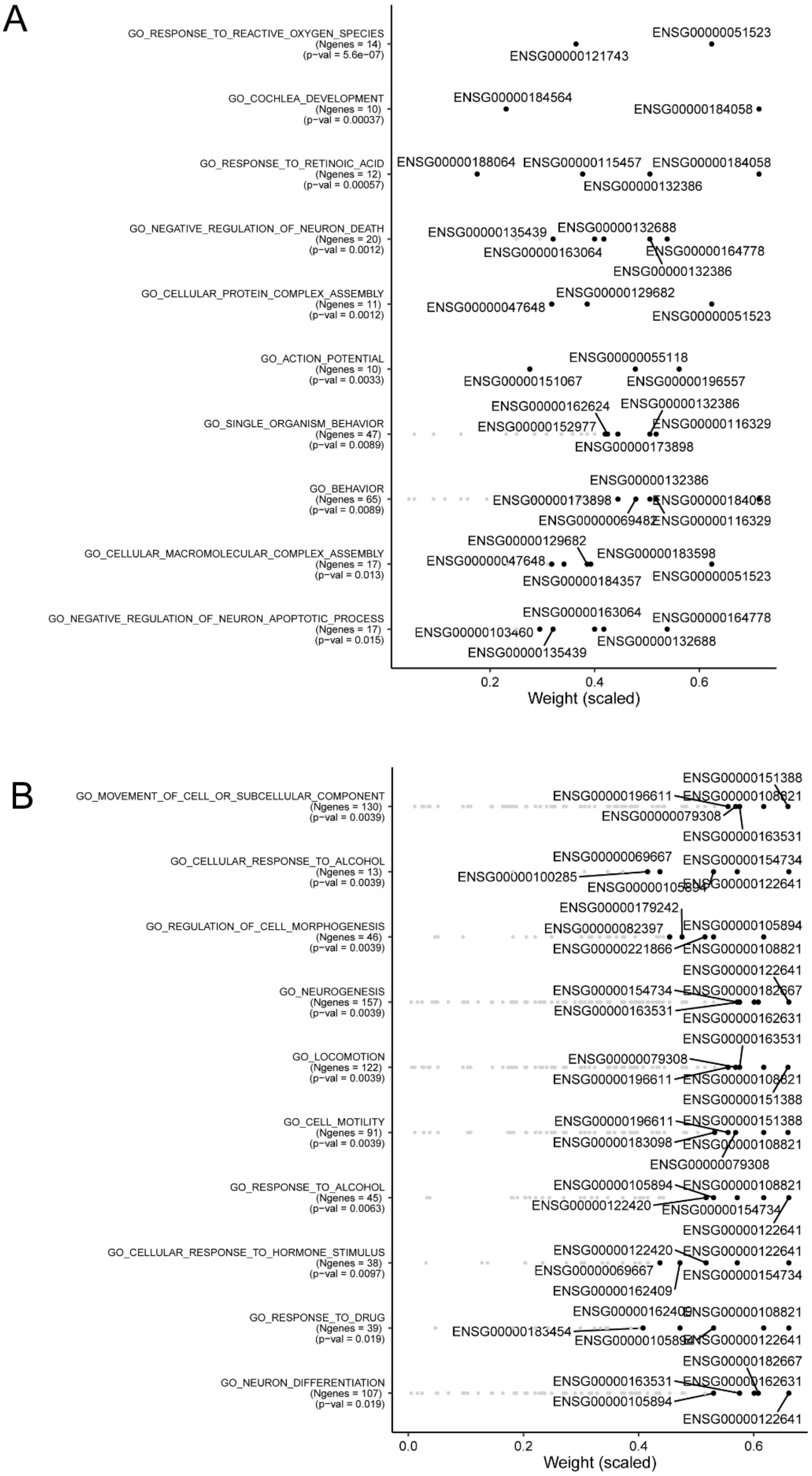
Supplemental Table 3. All differentially expressed miRNAs when comparing ALT+ and ALT- cell lines. Bold rows denote miRNAs with the GO-term "negative regulation of telomeres".

miRNA ID	baseMean	log2FoldChange	lfcSE	stat	p-value	padj
hsa-miR-485-5p	**103.7805395**	**23.74752621**	**2.091117**	**11.07663**	**1.63E-28**	**1.73E-25**
hsa-miR-708-5p	**3892.689986**	**7.264947052**	**0.678781**	**9.841096**	**7.49E-23**	**3.97E-20**
hsa-miR-767-5p	**542.2954246**	**8.04462707**	**0.819773**	**9.099622**	**9.06E-20**	**3.21E-17**
hsa-miR-105-5p	1051.810962	7.712397692	0.786521	9.061934	1.28E-19	3.40E-17
hsa-miR-377-5p	**59.41758256**	**24.20149455**	**2.863699**	**8.246848**	**1.63E-16**	**3.45E-14**
hsa-miR-3167	35.3434349	6.36090216	0.755065	7.649541	2.02E-14	3.57E-12
hsa-miR-708-3p	1334.986645	5.830520544	0.731637	7.169567	7.52E-13	1.14E-10
hsa-miR-4652-5p	7.967968373	22.36014243	3.185574	6.835547	8.17E-12	1.08E-09
hsa-miR-105-3p	23.03863142	6.761829679	0.96956	6.370753	1.88E-10	2.22E-08

hsa-miR-5579-3p	16.07228189	6.422509978	0.940723	6.205345	5.46E-10	5.79E-08
hsa-miR-767-3p	30.89226155	7.344326101	1.094828	6.173868	6.66E-10	6.43E-08
hsa-miR-204-5p	276.4130963	5.630930794	0.828096	6.093411	1.11E-09	9.77E-08
hsa-miR-3120-5p	238.2050991	-6.562024071	0.993275	-6.01749	1.77E-09	1.34E-07
hsa-miR-199b-5p	440.0676537	-4.807150735	0.715253	-5.90302	3.57E-09	2.52E-07
hsa-miR-139-5p	**1279.712325**	**4.546044042**	**0.72283**	**5.479909**	**4.26E-08**	**2.81E-06**
hsa-miR-214-3p	689.1759457	-6.108409186	1.00975	-5.47008	4.50E-08	2.81E-06
hsa-miR-5579-5p	7.861152725	5.720317536	0.963698	5.32876	9.89E-08	5.83E-06
hsa-miR-6716-3p	83.79944333	-3.047082809	0.465777	-5.28597	1.25E-07	6.98E-06
hsa-miR-138-5p	290.0261065	3.141467768	0.486688	5.252788	1.50E-07	7.95E-06
hsa-miR-520f-5p	83.68967803	6.63399073	1.188198	5.090895	3.56E-07	1.80E-05
hsa-miR-522-3p	47.87319285	6.881653716	1.247065	5.049179	4.44E-07	2.14E-05
hsa-miR-138-1-3p	25.59621833	3.354545841	0.550481	5.031137	4.88E-07	2.25E-05
hsa-miR-129-2-3p	**295.5711586**	**6.095149395**	**1.120096**	**4.919355**	**8.68E-07**	**3.84E-05**
hsa-miR-519d-5p	141.6618525	6.40855763	1.193808	4.878136	1.07E-06	4.55E-05
hsa-miR-129-5p	**14575.41183**	**5.390471303**	**1.004527**	**4.783814**	**1.72E-06**	**6.76E-05**
hsa-miR-139-3p	53.89802208	4.138250282	0.755176	4.705195	2.54E-06	9.61E-05
hsa-miR-3909	329.1472688	1.906169642	0.28847	4.579927	4.65E-06	0.00017
hsa-miR-204-3p	11.29913537	5.585052046	1.122078	4.456064	8.35E-06	0.000295
hsa-miR-490-3p	**25.54226318**	**6.510750486**	**1.337742**	**4.429666**	**9.44E-06**	**0.000323**
hsa-miR-129-1-3p	**29.69725573**	**6.230345756**	**1.298049**	**4.3491**	**1.37E-05**	**0.000453**
hsa-miR-214-5p	1164.062053	-4.199604413	0.841516	-4.29535	1.74E-05	0.000561
hsa-miR-142-3p	**23.86603606**	**3.877506414**	**0.770555**	**4.272901**	**1.93E-05**	**0.000602**
hsa-miR-3120-3p	**89.95014726**	**-3.525383223**	**0.689941**	**-4.26179**	**2.03E-05**	**0.000615**
hsa-miR-372-3p	**25.4339154**	**5.883600603**	**1.25354**	**4.22691**	**2.37E-05**	**0.000698**
hsa-miR-3117-3p	12.10842629	6.381595492	1.400878	4.137831	3.51E-05	0.000954
hsa-miR-323b-3p	655.3639747	6.425428894	1.409069	4.144886	3.40E-05	0.000954
hsa-miR-519a-3p	**14.05638776**	**6.059578824**	**1.322759**	**4.138757**	**3.49E-05**	**0.000954**
hsa-miR-218-5p	**9226.277131**	**3.030621899**	**0.595681**	**4.10559**	**4.03E-05**	**0.001044**
hsa-miR-34a-5p	1344.972613	3.651937015	0.746231	4.109903	3.96E-05	0.001044
hsa-miR-519a-2-5p	7.717589898	5.875451556	1.297878	4.07623	4.58E-05	0.001129
hsa-miR-520b-5p	7.717589898	5.875451556	1.297878	4.07623	4.58E-05	0.001129
hsa-miR-512-3p	**10.61673812**	**5.210568832**	**1.142544**	**4.048484**	**5.16E-05**	**0.001243**
hsa-miR-323a-3p	1022.523981	5.90847078	1.326748	4.012421	6.01E-05	0.001417
hsa-miR-122-5p	**75.80511244**	**2.844168008**	**0.570747**	**3.958264**	**7.55E-05**	**0.001741**
hsa-miR-146a-3p	29.32171218	6.034233386	1.437352	3.79116	0.00015	0.003385
hsa-miR-516a-5p	45.95966578	6.056353514	1.445981	3.783835	0.000154	0.003414
hsa-miR-10399-5p	18.54781669	-2.644159104	0.550346	-3.74157	0.000183	0.00396
hsa-miR-1284	9.080007248	2.984079988	0.644855	3.720342	0.000199	0.004222
hsa-miR-876-3p	6.725854433	4.030120611	0.930618	3.701969	0.000214	0.004451
hsa-miR-122b-5p	**18.89013071**	**2.98962442**	**0.687372**	**3.498289**	**0.000468**	**0.0092**
hsa-miR-375-3p	**194.1498475**	**3.835935052**	**0.928549**	**3.501092**	**0.000463**	**0.0092**
hsa-miR-4677-5p	9.097388935	-1.785090372	0.34305	-3.49829	0.000468	0.0092
hsa-miR-4662a-5p	6.642004326	4.251717395	1.054373	3.477629	0.000506	0.009759

baseMean , average normalized count values which have in turn been normalized to library size factor; Log2FoldChange, average fold change expressed in Log2; p-value, non-adjusted average p-value for gene; padj, average Benjamini-Hochberg adjusted p-value.

IV. Gene ontology analysis of genes in ALT multi-omics factor analysis signature



Supplemental Figure 1. Gene weights for gene ontology (GO) terms in ALT signature. The top two genes per GO-term are shown, with the rest in gray. (A) GO terms negatively weighted for ALT+ samples. (B) GO-terms positively weighted for ALT+ samples. N genes denotes the total amount of found genes in the dataset, p-val is adjusted p-value.