

<https://doi.org/10.1038/s42003-025-09005-8>

Genomic characterization of host gene alterations in *Theileria annulata*-transformed leukocytes



Debabrata Dandasena^{1,2}, Akash Suresh¹, Roli Budhwar³, Jeffrey Godwin³, Sakshi Singh^{1,2}, Madhusmita Subudhi¹, Vengatachala Moorthy A¹, Amruthanjali T¹, Sonti Roy¹, Vasundhra Bhandari^{1,4} & Paresh Sharma¹

Theileria annulata-infected host leukocytes exhibit cancer-like phenotypes, driven by mechanisms that remain incompletely understood. This study explores the genomic alterations underlying these transformations using whole-genome sequencing and bioinformatic analyses of six clinically relevant *T. annulata*-infected cell lines. Here we identify 7867 exon-linked somatic mutations shared across all cell lines, with significant enrichment in oncogenes (e.g., *FLT4*, *NOTCH2*, *MAP3K1*, *DAXX*, *FCGR2B*, *ROS1*) and tumor suppressor genes (e.g., *BARD1*, *KMT2C*, *GRIN2A*, *BAP1*). These mutations are associated with critical cancer-related pathways. Functional studies revealed that inhibition of the mutated oncogene *ROS1* using crizotinib induces death in infected leukocytes, confirming its role in transformation. Additionally, we observe mutations in genes linked to genomic instability and the DNA damage response (DDR) pathways, highlighting potential parallels with cancer biology. Suppression of TP53, a key tumor suppressor, is implicated in the immortalization of infected cells, while upregulation of the DNA mutator enzyme APOBEC3H suggests a parasite-driven, mutation-inducing mechanism. Our findings provide new insights into how *T. annulata* reprograms host cells through genomic instability and mutations, identifying *ROS1* and *TP53* as critical targets for therapeutic intervention. This work advances understanding of parasite-induced oncogenic transformation and offers pathways for future research.

Parasites in the phylum Apicomplexa cause significant morbidity and mortality in humans and animals worldwide. Among these, *Theileria annulata* and *Theileria parva* are the only parasites shown to trigger host cell transformation, a process with many of the hallmarks of cancer^{1–4}. Sporozoites of *T. annulata* is capable of infiltrating lymphocytes and macrophages, leading to the development of macroschizonts which can induce a malignancy in the host cells¹. The development of cancer-like characteristics in *Theileria*-infected leukocytes is associated with epigenetic modifications and changes in host gene expression^{1,4–6}. However, the precise mechanism through which parasite-infected leukocytes develop the cancer-like features remains unknown. Since *Theileria*-infected leukocytes share many characteristics with cancer, a genetic disease, it's plausible that somatic mutations in host genes might offer a selective advantage in suppressing apoptosis and inducing uncontrolled host cell proliferation¹.

Among the numerous potential causes of cancer, pathogen-induced DNA damage contributes to genomic instability and mutations in about 20% of cancers⁷. To survive and thrive, pathogens may directly or indirectly cause host cell damage that impairs their genomic integrity⁷. Loss of integrity can lead to genomic instability, a hallmark of cancer cells mainly driven by mutations in DNA repair genes, or faults in DNA replication genes^{8,9}. Both cancer and *Theileria*-transformed leukocytes are associated with uncontrolled cell growth, which may lead to genomic instability and an increased propensity for mutation acquisition. Infection of leukocytes by *Theileria* parasites leads to upregulated expression of some host genes such as *MDM2* (mouse double minute 2 homologs; *p53* negative regulator) and *SMYD3* (SET and MYND domain containing 3), which may induce genomic instability in infected cells, although no direct relationship has been demonstrated^{7,10}. Besides *Theileria*, *Cryptosporidium*, another apicomplexan parasite, is also

¹National Institute of Animal Biotechnology, Hyderabad, India. ²Graduate Studies, Regional Centre for Biotechnology (RCB), Faridabad, India. ³Bionivid Technology Pvt. Ltd., Bangalore, India. ⁴National Institute of Pharmaceutical Education and Research (NIPER), Hyderabad, India. ✉e-mail: paresh@niab.org.in

recognized for inducing cancer-like symptoms in experimental models through disruption of host DNA integrity^{1,7,11}.

Because cancer is a genetic disease, mutations are extensively studied for any connection to cancer, but their potential significance in *Theileria*-infected leukocytes remains unclear¹². Cancer-like characteristics, such as genomic instability and acquired mutations, have not yet been explored in *Theileria*-transformed leukocytes. Genomic instability and mutation in *Theileria*-infected leukocytes should be examined to see whether infection and resulting cellular transformation provoke instability and/or changes to the host genome, and if so, does parasite infection cause cancer in host cells?

Next-generation sequencing has made it possible to systematically discover the mutational spectra of various cancers that may arise from inherited mutations, environmental causes, or faulty DNA replication¹³. In this study, we used whole-genome sequencing (WGS) to examine how *T. annulata* infection affects host genome integrity by profiling somatic mutations in infected cells. We hypothesized that *Theileria*-infected leukocytes may acquire a selective advantage through mutations in oncogenes or tumor suppressor genes (TSGs), thereby contributing to the mechanisms of parasite-mediated host cell transformation. We identify linkages that enhance our knowledge of the *Theileria* parasite's adaptation to its host leukocyte, a significant portion of which has not been previously investigated. Identifying these traits opens the path for creating and using cancer therapeutics to treat tropical theileriosis caused by *T. annulata* and a greater understanding of the molecular connections underlying the resulting cancer-like phenotypes.

Results

Whole genome characterization of somatic mutations in lymphoproliferative cells infected with *T. annulata* parasites

To study the impact of *T. annulata* infection on host genome integrity, we conducted WGS analysis on six transformed leukocyte cell lines. Confirmation of parasites in all cell lines was made by utilizing PCR and IFA, which were directed towards TASP, a gene exclusive to *T. annulata* parasites. An IFA image demonstrating the presence of *T. annulata* in these infected cell lines is shown (Supplementary Fig. 1). The six cell lines have been previously characterized using microsatellite markers and genotype-based sequencing analysis¹⁴. We analyzed the sequenced samples using the *Bos taurus* genome (ARS UCD 1.2), as a reference to identify changes in cancer-related genes. Supplementary Data 1 shows the number of reads per sample, and a flowchart of the overall WGS statistics plus workflow is shown (Supplementary Fig. 2). Variants were evaluated only if they met the following criteria: a Phred quality score of at least 30, an allele frequency of at least 8%, and a significance level of $P < 0.05$. Variants with a variant allele frequency of $>60\%$ were called homozygous. Furthermore, we excluded non-coding data concentrating on coding regions to identify synonymous, and non-synonymous (missense) mutations. In recent years it has become clear that synonymous mutations may also contribute to cancer and are no longer considered neutral^{15,16}. Each sample had an average of 98,000 somatic mutations (SNPs) and 3400 indels in coding regions (Fig. 1a, b). We next looked for alterations common to all samples to identify genes that might play a role in inducing the cancer-like phenotype of *Theileria*-infected leukocytes. A total of 7867 shared mutations were split between non-synonymous ($n = 3590$) and synonymous ($n = 4262$) variants, representing 3580 genes (Supplementary Data 1.xls). There were 1136 cases of homozygous and 6731 cases of heterozygosity. Besides the single gene variations, we also detected 412 common indels in genes across the WGS data sets (Fig. 1b).

Furthermore, we used the TCGA and COSMIC databases to identify 127 genes that may be implicated in the cancer-like phenotype of *Theileria*-infected leukocytes. Comparative analysis identified significant mutations in the genes known to have an important role in cancer. The TCGA dataset provides a comprehensive list of genes that are known to participate in various cellular functions across multiple types of cancers¹⁷. By analysing our own data sets, we have identified 127 distinct significantly mutated genes (SMGs), indicating the potential involvement of these cellular and

enzymatic mechanisms in *Theileria*-induced leukocyte transformation. As per TCGA dataset, these SMGs play crucial roles in essential cellular processes associated with cancer, including cell cycle regulation, mitogen activated protein kinase (MAPK) signaling, phosphatidylinositol-3-OH kinase (PI(3)K) signaling, Wnt/ β -catenin signaling, transcription factors/regulators, epigenetic regulation, maintenance of genome integrity, and ubiquitin-mediated proteolysis. (Fig. 1c).

Next, using the COSMIC database, we searched for mutations in cancer hallmark genes involved in oncogenic processes such as proliferation, growth suppression, replicative cell immortality, invasion and metastasis, genomic instability and mutations, evasion of programmed cell death, and changes in cellular energy metabolism. Except for genomic instability and mutations, these cancer-like characteristics are documented in the context of the *Theileria*-infected leukocytes¹. We discovered 121 mutated genes common to all 6 samples, 89 of which were ranked as “tier-1” and 32 as “tier-2” in the COSMIC database. Since mutations in “tier-1” genes have been directly shown to promote cancer, we next identified 22 genes with non-synonymous mutations (Supplementary Data 1.xls). In addition to the common missense mutations in the 22 COSMIC tier-1 genes, each gene also possessed unique alterations that may be functionally significant for its encoded activity. The bar graph depicts the mutational load of COSMIC tier-1 genes ($n = 22$) in *Theileria*-infected leukocytes (Fig. 2a). Twelve of these tier-1 genes are also cancer hallmark genes, as mutations in them have been demonstrated to have a direct role in cancer progression: *FLT4*, *NOTCH2*, *MAP3K1*, *DAXX*, *BARD1*, *KMT2C*, *GRIN2A*, *BAP1*, *FCGR2B*, *ROS1*, *SLC34A2*, and *NOTCH1* (Fig. 2b). Using the SIFT score, we examined the influence of non-synonymous mutations on the cosmic listed genes in our datasets. We discovered mutations in *KMT2C*, *TSC2*, *ROS1*, *HOOK3*, *LZTR1*, *CHEK2*, *TAF15*, and *GNA11* that might impact their function. *KMT2C* and *ROS1* were discovered to have the highest number of mutations in our samples. Fig. 2c–d shows a lollipop plot for *KMT2C* and *ROS1*, illustrating the distribution of variants and their effect on the genes. To assess the impact of mutations on specific candidate tumor suppressor genes (*BARD1*, *KMT2C*, *GRIN2A*, *SLC34A2*, *NOTCH1*, and *BAP1*), we conducted q-PCR analysis on cell lines infected with *Theileria*, using healthy PBMCs as a control. Our findings revealed a significant decrease in the expression of all the examined tumor suppressor genes in the parasite-infected lines. These mutations may lead to the absence or impaired functioning of the proteins encoded by these genes, potentially resulting in uncontrolled cell division and contributing to the development of cancer-like phenotypes in the parasite-infected cells (Fig. 2e). However, further studies would be required to prove mutation-driven expression changes.

Alterations in genes associated with the genomic instability pathway, such as *DAXX*, *BARD1*, *KMT2C*, and *BAP1*, may explain the increased frequency of mutations in *Theileria*-infected cells. A defect in the DNA repair pathways increases the risk of cancer and genomic instability¹⁸. We discovered common mutations in 58 genes involved in important DNA repair pathways across our datasets (Supplementary Table 2.pdf). The major DDR-related genes and pathways that have been mutated are homologous recombination (*MUS81*, *XRCC2*, *PALB2*, *RPA2*, *EME1*), non-homologous end joining (*PRKDC*, *MAD2L2*), mismatch repair (*PMS1*), nucleotide excision repair (*GTF2H4*, *LIG1*, *UVSSA*, *ERCC6*, *GTF2H3*, *CCNH*), base excision repair (*NEIL2*), fanconi anaemia repair (*FANCD2*, *FANCM*, *FANCC*), DNA polymerase (*POL1*, *POLN*, *POLG*, *POLD1*, *POLE*), ubiquitin and modification (*RNF8*), repair of DNA-protein crosslinks (*TDP1*), and other conserved DDR pathways proteins like *CLK2*, *CHEK2*, *TOPBP1* and *MDC1* (Fig. 3a). *CHEK2*, a TSG gene generally engaged in DNA repair, had both shared and distinct synonymous and non-synonymous mutations across the samples (Fig. 3b). The detected mutations in the *CHEK2* gene have a deleterious effect (SIFT < 0.05) on the gene's normal function, indicating a likely DNA repair pathway defect in *Theileria* infected cells.

In addition to cancer-related gene mutations, we found common homozygous non-synonymous alterations in epigenetic regulators such as *KMT2C*, *ARMT1*, *TRMT1L*, *TRMT61A*, *EHMT2*, *METTL27*, *ALKBH4*,

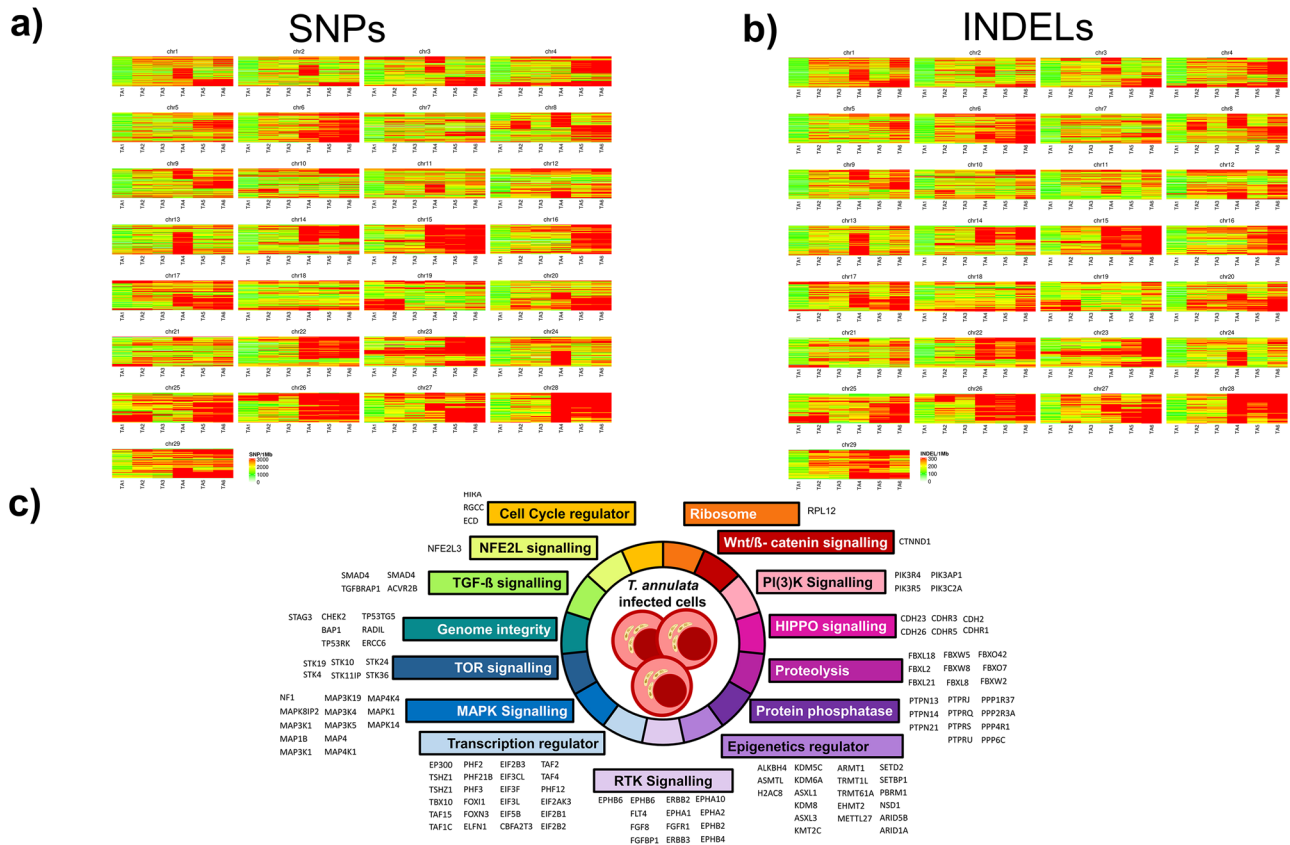


Fig. 1 | Chromosome-wise distribution of SNPs, INDELs, and significantly altered pathways in *T. annulata* infected samples. a The heat map represents the chromosome-wise distribution of all the SNPs across 29 chromosomes for 6 WGS samples used in the study. **b** Heat map representing the chromosome-wise

distribution of all the Indels for 6 WGS samples used in study. **c** Cancer enrichment analysis using TCGA database showing major signaling pathways altered across our samples based on the gene mutations. The figure shows pathways and related genes mutated across 6 *Theileria* infected cells identified using WGS analysis.

ASMTL, H2AC8 and EHMT2 (Supplementary Data 1.xls). Since epigenome regulator mutations are possible therapeutic targets in cancer, they might be crucial in the life cycle of *T. annulata* and beneficial in developing therapies against the parasite¹⁹.

ROS1 contributes to *T. annulata*-infected leukocyte proliferation WGS of *T. annulata*-infected leukocytes identified *ROS1* as the most frequently mutated tier-1 cancer gene, exhibiting the highest mutational burden among analyzed oncogenic loci. As *T. annulata*-infected leukocytes continuously proliferate it appears that this high mutational burden does not negatively impact on infected leukocyte proliferation in vitro. Importantly, nor did it negatively impact on *ROS1* expression, as cytoplasmic *ROS1* can be readily detected in infected cells (Fig. 4a). *Theileria* infection appears upregulate expression of *ROS1* compared to uninfected host PBMCs, where *ROS1* was undetectable (Fig. 4b).

Crizotinib is a tyrosine kinase inhibitor approved by the US Food and Drug Administration (FDA) for the treatment of certain types of advanced non-small cell lung cancer (NSCLC) harboring *ROS1* mutations²⁰. The contribution of upregulated *ROS1* to infected leukocyte proliferation was therefore examined using crizotinib. Inhibitor treatment revealed dose-dependent anti-proliferative effect of crizotinib, with IC₅₀ values of 1–15 μM across infected cell lines (Fig. 4c). Time-course experiments at IC₅₀ concentrations showed marked reduction in cell viability (Fig. 4d) and concomitant downregulation of *ROS1* protein (Fig. 4e). To gain support for *ROS1* being the principal target of crizotinib in *Theileria*-transformed leukocytes transcriptional profiling was performed. This demonstrated significant *ROS1* overexpression in infected cells, while ALK and c-MET (also known to be crizotinib targets) were downregulated relative to control PBMCs (Fig. 4f, g). Taken together, the significant upregulation of *ROS1*

compared to ALK and c-MET and crizotinib's anti-proliferative effect strongly implicates *ROS1*-dependent signaling as a mediator of *Theileria*-infected leukocyte proliferation. Clearly, deeper investigation is merited to determine just how *ROS1*-dependent signaling affects proliferation.

DNA mismatch repair (MMR) related signatures lead to Microsatellite instability (MSI) in *Theileria*-infected cells

WGS has advanced cancer genetics to the point that cancer-type-specific mutational signatures have been discovered²¹. These mutational fingerprints represent endogenous and exogenous cancer factors^{22,23}.

Multiple mutational mechanisms cause somatic mutations in cancer genomes, each leaving a mark. Mutational signature analysis may assist in determining if there is a relationship between observed mutations and genomic instabilities, or defective DNA repair pathways. Since mutational signatures for most cancer types are readily accessible, this analysis would also help us determine whether *Theileria*-transformed leukocytes and human cancer share a common signature. Based on the orientation of mutations, all point substitutions were classified into six types (C > A, C > G, C > T, T > A, T > C, T > G) using SomaticSignatures (Fig. 5a–c). The proportions of all six mutation groups were comparable across samples, with C > T and T > C being the most prevalent mutation types based on the sequencing data (Supplementary Data 2.xls). Using deconstructSigs we next established if any mutational signatures are associated with specific mutation subtypes. It revealed signatures characteristic of SBS1A, SBS20, and SBS12 in all six infected leukocyte lines (Fig. 5d). SBS1A (associated with deamination of 5-methyl cytosine) signatures are ubiquitous among different cancer types and present in most cancer cell lines²². SBS20 is associated with concurrent *POLD1* mutations, defective DNA mismatch repair (MMR), and microsatellite instability (MSI)²⁴. *POLD1* mutation was present

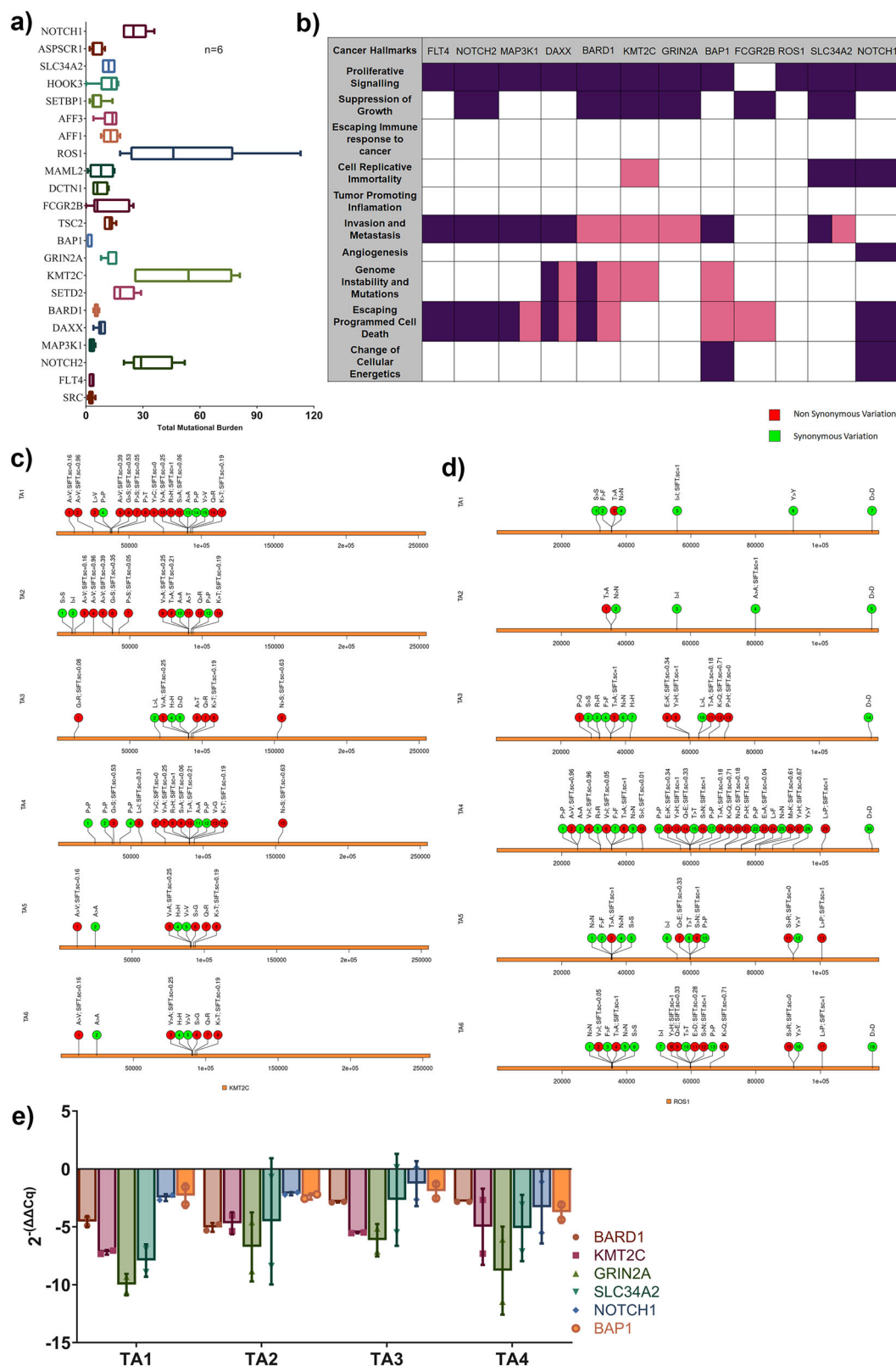


Fig. 2 | Mutation in the cancer hallmark genes and their functional relevance in *T. annulata* infected samples. **a** The graph shows the total mutational burden of the 22 tier 1 gene (COSMIC) identified in our data using WGS analysis. Gene's mutational burden was calculated individually in each sample and then averaged for the six samples. Both synonymous and non-synonymous mutations were calculated for all six samples and represented in the form of a box plot as the total mutational burden. Error bars represent the mean with standard deviation (SD). **b** Based on the COSMIC database, the Figure shows the list of cancer hallmark genes and their process

linked. These 12 cancer hallmark genes are found to have common nonsynonymous mutations throughout the six samples in our data. These genes are shown to have roles as tumor activators or suppressors marked with purple and pink colors, respectively. **c, d** Lollipop images representing the distribution of mutations along with the SIFT score of KMT2C and ROS1 across six *Theileria*-infected cells. **e** q-PCR based gene expression levels of tumor suppressing genes in *T. annulata* cells w.r.t uninfected healthy PBMCs. Error bars represent the mean with standard deviation (SD).

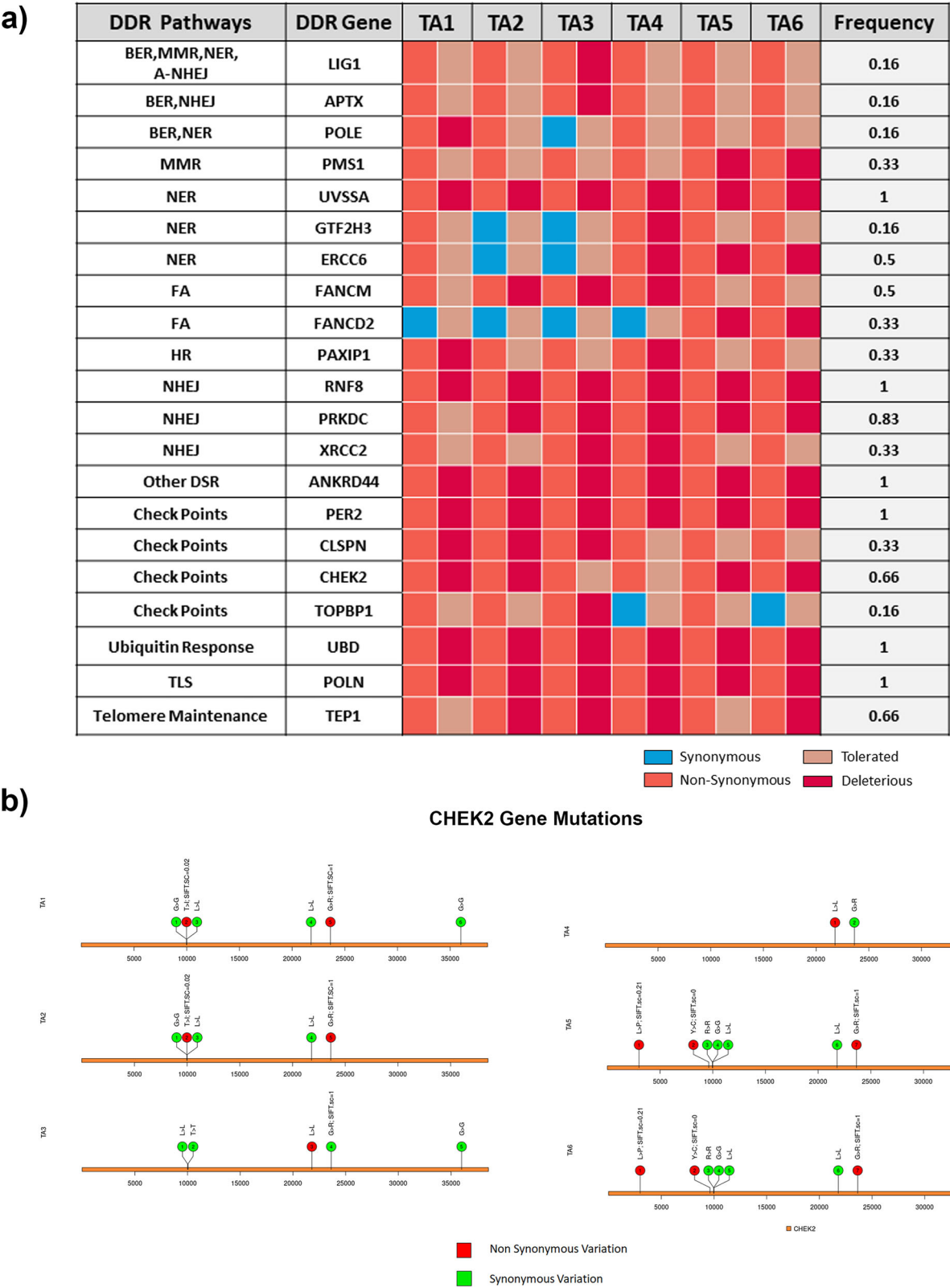
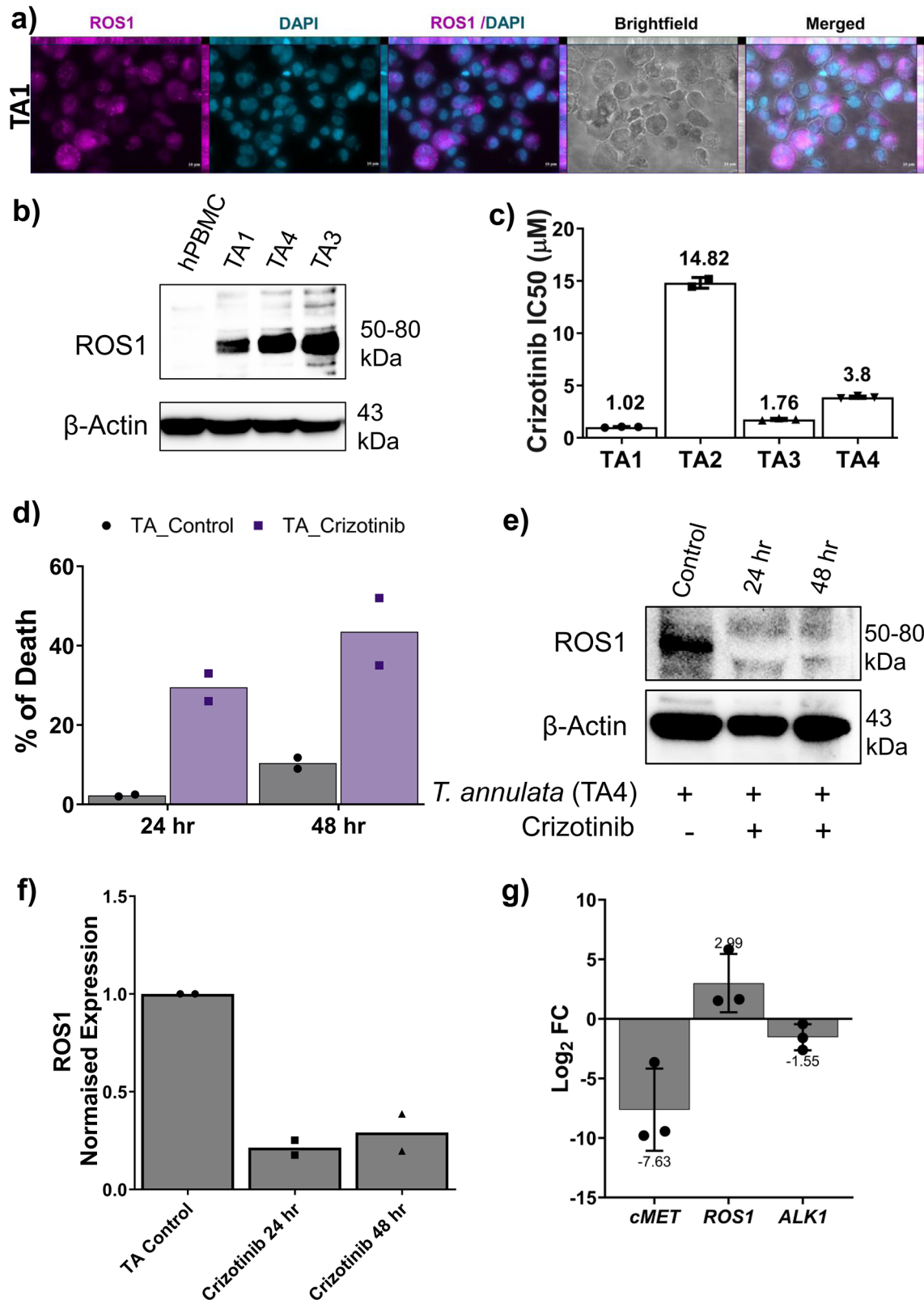


Fig. 3 | Mutational spectrum of DDR genes in the *T. annulata* infected samples. **a** The Figure shows major DDR pathway-related genes altered in our samples and the frequency of their alteration in the six datasets. Synonymous mutations are marked with blue and non-synonymous with orange color. The impact of the mutations was marked with another two sets of colors, light orange represents the tolerated mutation, and red represents the deleterious mutations predicted from the SIFT score. Samples lacking any mutation for the genes are marked with gray color. *Represent the presence of missense variation. The frequency of each gene is represented at the end. **b** Lollipop image showing the mutations of the CHEK2 gene for all six samples along with the SIFT score. All six samples have multiple mutations representing non-synonymous (red) and synonymous(green). Four out of the six samples had deleterious mutations, marked by the SIFT score.



across all of 6 of the infected leukocyte cell lines, validating the results of the mutational signature (Supplementary Data 1.xls). Mutations in DNA mismatch repair genes (MMR) induce microsatellite instability (MSI), a hypermutable trait that increases cancer risk²⁵. Four genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* govern the fate of the MMR pathway. We next asked if any mutations were present in the four MMR genes in *Theileria*-infected

leukocytes. Both *MSH2* and *PMS2* were mutated in all six *Theileria*-infected leukocyte cell lines, while *MSH6* was mutated in all but one line. In addition, mutations were also found in additional MMR genes (*MSH3*, *MSH4*, *MSH5*, *PMS1*, *HFM1*, *MSX2*) in five of the *Theileria*-infected leukocyte cell lines (Supplementary Data 2.xls). As MMR may result in microsatellite instability (MSI), we next examined the *Theileria*-infected leukocyte cell lines for the

Fig. 4 | Impact of Crizotinib on ROS1 Expression and Cell Viability in *T. annulata*-Infected Cells. **a** Immunofluorescence analysis showing ROS1 expression in TA1 cells. Cells were stained with anti-ROS1 (magenta) and DAPI for nuclear staining (blue). Brightfield image is included for reference. The presented image is a maximum intensity projection (MIP) of a Z-stack acquired using fluorescence microscopy. **b** Western blot analysis showing ROS1 expression (50–80 kDa) in healthy peripheral blood mononuclear cells (hPBMC) and *T. annulata*-infected cell lines (TA1, TA3, and TA4). β -actin (43 kDa) was used as a loading control. The blot shown is representative of three independent experiments ($n = 3$). **c** Bar graph representing the IC50 values of crizotinib in four *T. annulata*-infected cell lines (TA1, TA2, TA3, and TA4). Data are derived from three biological replicates, and error bars indicate the standard error of the mean (SEM). The mean IC50 value for each cell line is displayed above the corresponding bar. **d** Bar graph representing the percentage of cell death in *T. annulata*-infected cells following crizotinib treatment

at 24 and 48 h. Untreated controls for each time point are included for comparison. Statistical analysis was performed between treated and corresponding control groups at each time point. Data are presented as mean \pm SEM from 3 independent experiments. **e** Western blot analysis of ROS1 expression in the TA cell line following crizotinib treatment for 24 and 48 hours. Cell lysates from treated and untreated (control) samples were collected and probed for ROS1. β -actin was used as a loading control. **f** Densitometric quantification of ROS1 expression from two biological replicates, normalized to β -actin. Data are represented as a bar graph comparing treated samples to control at each time point. **g** qPCR analysis of ROS1, cMET, and ALK1 expression in *T. annulata* infected cells. Data are represented as log2 fold change (FC) relative to healthy PBMCs, which were used as the non-infected control. The experiments were performed with three biological replicates, and HPRT was used as the housekeeping control.

presence of MSI (Supplementary Data 3.xls). Using the WGS datasets, we measured the lengths of specific DNA microsatellites present in *Theileria*-infected leukocytes (Fig. 5e). MSI was detected in all 6 infected leukocyte cell lines, with a mean MSI index of 24.19 ± 3.38 per sample (Fig. 5f).

Copy number variation (CNV) analysis identified alterations in cancer-related genes in the *Theileria*-infected cells

CNVs influence more of the genome than other somatic mutation in cancer cells, activating oncogenes and inactivating TSGs²⁶. Therefore, we asked whether there were any CNVs associated with the cancer-like phenotype of *Theileria*-infected and transformed leukocytes. We exploited CNVcaller to identify the CNVR (copy number variation regions) across the six *Theileria*-infected leukocyte cell lines ($n = 6$) (Fig. 6a–d) (Supplementary Data 4.xls).

Next, we searched for CNVs in the cancer-related genes using the 2013–08–2016 TCGA Pan-cancer dataset. The analysis unveiled amplification of 10 genes that are frequently reported as significantly altered in various types of cancers, within the *Theileria*-infected cell lines (Fig. 6e). The genes were: *NFE2L2*, *MPHOSPH8*, *MACROD2*, *BCL2L1*, *PARP10*, *MDM4*, *CCNE1*, *ERBB2*, *PAX8*, and *BPTF*. DNA alterations frequently occur in cancer and are known to drive aneuploidy and chromosomal instability^{27,28}. Given that normal cells typically struggle to tolerate such imbalances, we investigated whether *Theileria*-infected leukocytes exhibit changes in DNA content. Using DNA flow cytometry, we assessed ploidy levels in *T. annulata*-infected leukocytes ($n = 2$), with healthy cattle leukocytes serving as diploid controls.

Flow cytometric analysis detected pronounced ploidy deviations in infected leukocytes. The DNA index of 1.58 in these cells (Fig. 6f–g) surpassed the 1.1 aneuploidy threshold. Histograms displayed bimodal distributions, revealing both diploid ($DI = 1.0$) and aneuploid subpopulations, the latter showing elevated DNA content. These results align with whole-genome sequencing data, which identified widespread copy number variations across six independent datasets. The consistency between cytometric and genomic evidence strongly suggests that *Theileria* infection induces chromosomal instability in host leukocytes, leading to detectable aneuploid populations.

Different malignancies and cancer cell lines all share a significant association between CNV and altered gene expression²⁶. For independent confirmation of the CNV discovery-based approach q-PCR analysis of two candidate genes *BCL2L1*, and *NFE2L2* was performed. This revealed increased expression of the anti-apoptotic gene *BCL2L1* and the transcription factor *NFE2L2*, which both showed amplification in the CNV dataset (Fig. 6h). Enhanced mRNA expression of the *BCL2L1* and the presence of a CNV in the gene might be responsible for the resistance to apoptosis phenotype observed in *Theileria*-infected and transformed leukocytes¹.

Identification of large-scale Structural variations across *T. annulata* infected cells

Genomic deletions, duplications, and rearrangements may affect anything from a few kilobases to an entire chromosome, known as structural variation

(SVs), which plays a critical role in the development of cancer²⁹. Next, we wanted to discover similar genetic alterations in *Theileria*-infected cells since SVs play a vital role in changing gene expression and are one of the leading factors generating the cancer phenotype. We analysed the genomes of the *Theileria*-infected cells using the algorithm Breakdancer to identify the four major somatic structural variants: deletions (DEL), inversions (INV), and translocations (CTX/ITX). This algorithm helped us to detect huge SVs (>1Kb), but for further analysis, we only took SVs with a confidence score of 99 (Supplementary Data 5.xls). Fig. 7a–f illustrates the SVs discovered in each of the six-cell lines using the CIRCOS plot. On average, 2427.3 ± 575.29 variations were detected across all samples. Out of which, 17.32 ± 6.53 , 36.88 ± 10.81 , 15.84 ± 12.51 , and $29.93 \pm 9.03\%$ belong to CTX, DEL, INV, and ITX, respectively (Fig. 7g). Sixteen genes had common SVs across all the *Theileria*-infected cells; these include *SESTD1*, *ITPR2*, *ELMO1*, *IGF2BP2*, *HAT1*, *PPM1H*, *UBR4*, *DPP6*, *ZNF654*, *PDE5A*, *LNPI*, *GRID2*, *CGGBP1*, *RAP1A*, *CACNA1C*, *FABP2* (Fig. 7h). Our investigation of SVs led us to *RAP1A*, a protein belonging to the RAS oncogene family that controls signaling pathways affecting cell proliferation and adhesion and may play a role in tumor malignancy³⁰.

Nutlin-3a disruption of TP53 binding to MDM2 provokes *Theileria*-transformed leukocyte death

It is well-established that treatment infected leukocytes with buparvaquone (BPQ) induces parasite death and this results in loss of the immortalized phenotype characterized by TP53 activation, nuclear translocation, and subsequent host cell apoptosis³¹. TP53 activity depends on its binding to MDM2 so Nutlin-3a was employed to disrupts TP53-MDM2 binding^{6,32}. Nutlin-3a treatment recapitulated BPQ effects (Fig. 8a–c) demonstrating that pharmacological TP53 activation alone is sufficient to reverse the *Theileria*-transformed phenotype. Immunofluorescence analysis revealed nuclear accumulation of TP53 (green) following 48 h of Nutlin-3a exposure, while TaSP staining (magenta) confirmed persistent parasite presence until infected leukocytes started to die. Quantitative analysis showed significant TP53 upregulation (Fig. 8b) correlating with progressive loss of infected leukocyte viability 48–60 h post-treatment (Fig. 8c). Whole genome sequencing confirmed TP53 integrity in transformed leukocytes, with no SNP detected. These findings establish that: (1) viable parasites maintain transformation by provoking MDM2-mediated TP53 sequestration, (2) TP53 remains functionally competent in infected cells, and (3) reactivation of endogenous TP53 initiates leukocyte cell death in the presence of live parasites. Mechanistic parallels to cancer cell survival pathways³³ suggest strategies targeting TP53 offer therapeutic potential against *Theileria*-induced lymphoproliferation^{34,35}.

Cytidine deaminase APOBEC3H might contribute to cancer like mutagenesis in *Theileria*-infected cells

We next wanted to find the origin of somatic mutations, which we have discovered in our WGS analysis that might be helping the host cells to get immortalized. Recurrent infection with *Plasmodium* parasites has been shown to promote genomic instability and AID-Dependent B Cell

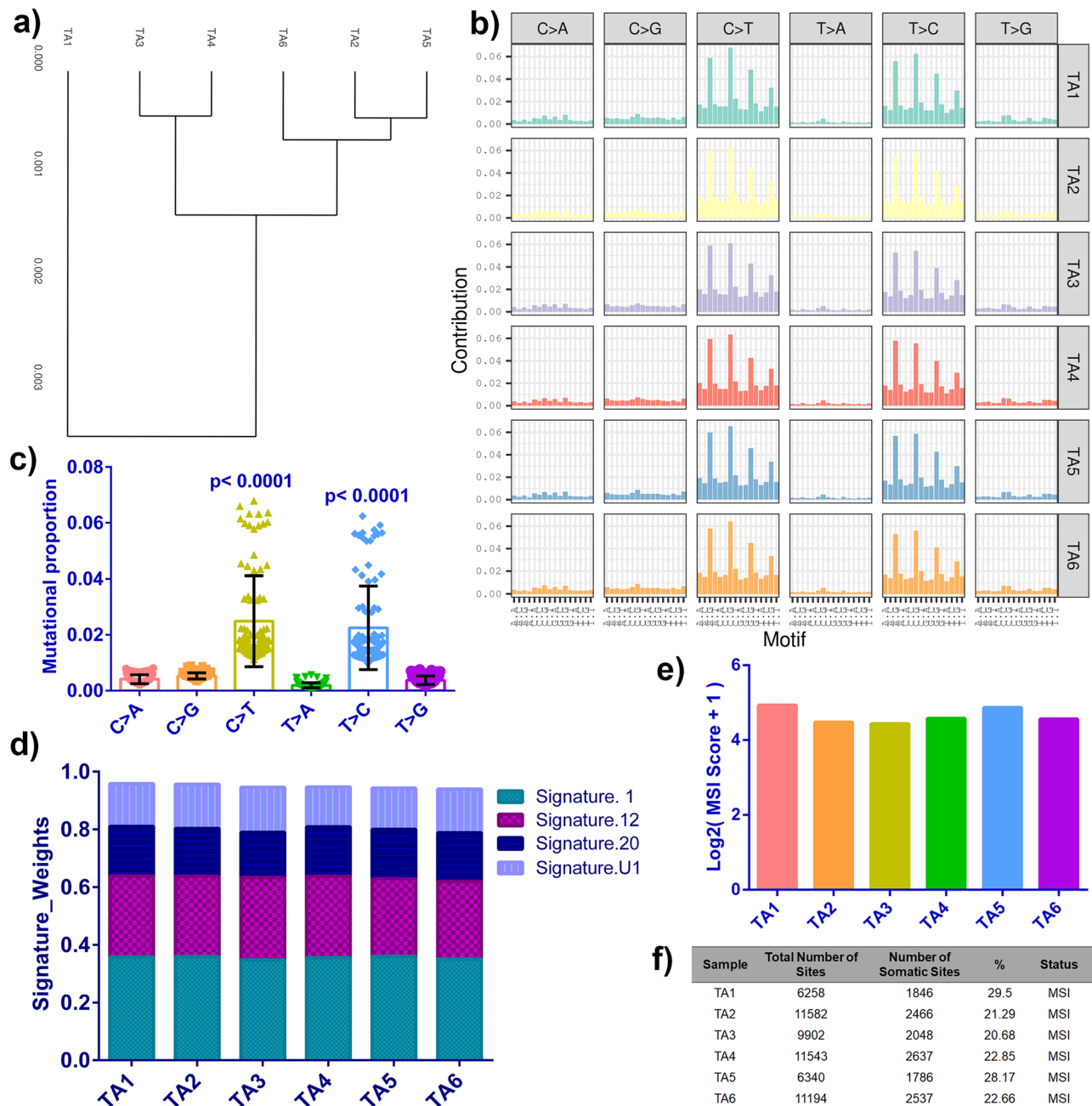


Fig. 5 | Divergent mutational features and MSI status of *T. annulata* infected bovine lymphocytes compared to the reference. **a** Dendrogram represents the hierarchical clustering of the mutational spectrum across six *T. annulata* infected samples. **b** Contribution of six mutational subtypes (C>A, C>G, C>T, T>A, T>C, T>G) represented in the form of a bar chart for all six different datasets (TA1-TA6), marked with different color codes. **c** C>T and T>C substitutions are significantly elevated across all six samples compared to the other four. Mutational proportion data of all the 16 possible tri codons for each substitution for all six samples have been represented as a scattered plot. The error bars reflect mean and standard deviations. One-way ANOVA analysis was carried out to perform multiple

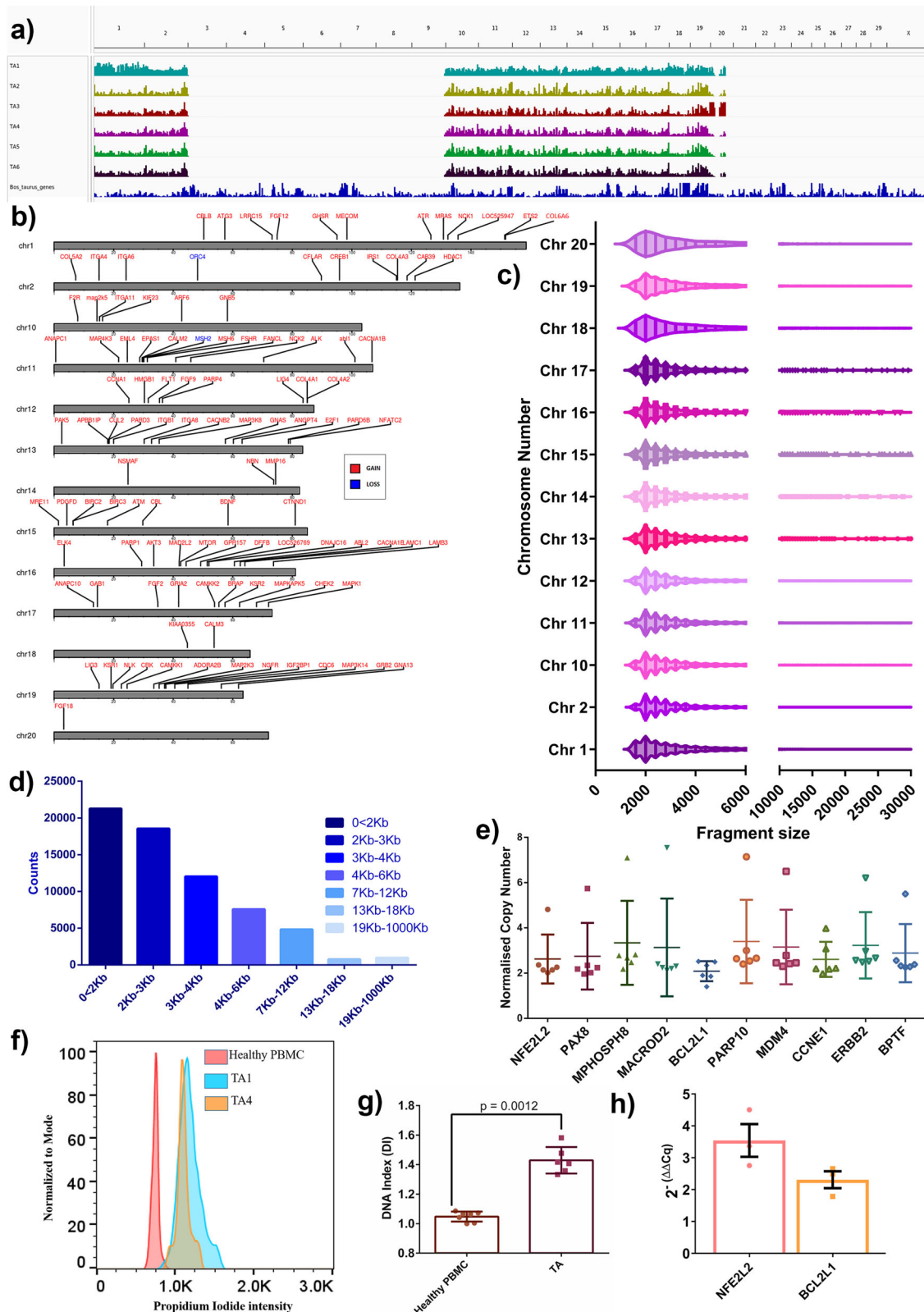
comparisons among all six single-base substitutions. The graph represents the frequency of each substitution for all six samples. Tukey's multiple comparisons test was used and the p -value was calculated **** ($p < 0.0001$). **d** Mutational signature present in all *T. annulata* infected samples: We reconstructed the proportion of mutational signatures of each sample based on a predefined mutational spectrum of 30 COSMIC signatures. The four identified signatures in the infected samples are shown. **e** The bar graph displays the MSI score of all six samples in the $\text{Log}_2(\text{MSI Score} + 1)$ scale. **f** The table represents the percentage of MSI sites along with the MSI status of all six samples.

Lymphoma³⁶. We next investigated the possibility of such events in the *Theileria*-infected cells by checking the expression of common DNA mutators like activation-induced cytidine deaminases (AID) family genes, specifically *AICDA* and *APOBEC* gene family members. We examined *APOBEC3A*, *APOBEC3B*, *APOBEC3H* and *AICDA* mRNA expression levels in our cell lines by qPCR (Fig. 8d). Within *Theileria*-infected cells, there was a substantial increase in *APOBEC3H* mRNA expression, while the expression of other genes exhibited a decline when compared to cells from

healthy animals. These findings indicate that *APOBEC3H* could potentially contribute to somatic mutations in our cells.

Discussion

This paper introduces the framework of somatic mutations observed in *T. annulata*-infected leukocyte cell lines. These mutations have the potential to induce genomic instability in infected leukocytes, thereby contributing to their transformation resembling cancer-like characteristics. The WGS data



were generated on infected leukocytes isolated from different animals harboring the same *T. annulata* parasite. This approach enables comparison of genomic alterations in infected cells, though we recognize that the absence of matched healthy controls from the same animals limits our ability to fully distinguish parasite-induced changes from pre-existing host genetic variation or environmental influences. To identify cancer-like somatic

mutations, we compared the data with publicly accessible datasets from either healthy animals or the COSMIC and TCGA databases. As the goal was to identify genetic links between *Theileria*-transformed leukocytes and cancer we restricted our attention mainly to mutations in gene coding regions, as they have been extensively documented for a role in carcinogenesis.

Fig. 6 | Analysis of CNVs of six *T. annulata* infected samples. **a** The graph illustrates the chromosomal distribution of copy number variations (CNVs) across six samples, using the *Bos taurus* genome as a reference. The bottom histogram represents the CNV profile of the reference genome, providing a baseline for comparison. The six individual plots above correspond to the CNV profiles of each sample, arranged chromosome-wise. Deviations from the baseline indicate copy number gains or losses, highlighting genomic regions with potential amplifications or deletions. **b** The figure displays key genes found to be commonly affected across all six samples, mapped chromosome-wise. Each horizontal line represents a chromosome, labeled with its corresponding number. Gene names in red indicate regions with copy number gains, while gene names in blue represent regions with copy number losses. **c** The horizontal bar graph represents the chromosome-wise distribution of fragment sizes for both amplifications and deletions. Error bars indicate the mean \pm standard deviation (SD), showing the variation across samples. **d** The bar graph represents the absolute count of copy number variations (CNVs) across different size categories. The x-axis denotes CNV size ranges (e.g., 0–2 kb, 2–3 kb), while the y-axis shows the total number of variations in each category. The number

of CNVs is highest in the smallest size range (0–2 kb) and gradually decreases with increasing fragment size. **e** The scatter plot presents the normalized copy number values for individual genes that were significantly altered across multiple cancer types, analyzed in all six samples. Each point represents a gene, with error bars showing the standard deviation (SD) across the samples. **f** The histogram represents the flow cytometry-based DNA content in *Theileria annulata*-infected cells (TA1 and TA4) compared to healthy PBMCs. The x-axis denotes propidium iodide intensity, while the y-axis shows the cell number normalized to mode. **g** The bar graph represents the DNA index derived from flow cytometry data, comparing healthy PBMCs and *Theileria annulata*-infected cells (TA). A Mann-Whitney test yielded a p-value of 0.0012, indicating a statistically significant difference. Bars represent the mean DNA index, with error bars showing the standard deviation (SD). **h** The bar graph represents qPCR-based relative gene expression analysis of NFE2L2 and BCL2L1, using HPRT as the housekeeping control. Fold change was calculated using the $2^{-\Delta\Delta CT}$ method, and error bars indicate the standard error of the mean (SEM).

The SNP analysis identified alterations in gene coding regions ($n = 7867$) common to six *Theileria*-infected leukocyte cell lines. Comparative analysis using the TCGA and COSMIC databases allowed the identification of genes that are often changed in most cancers. The analysis revealed mutations in genes implicated in several cancer-related cellular and enzymatic processes, as shown by 127 distinct SMGs in *Theileria*-infected leukocytes. Comparison with the COSMIC database detected mutations in “hallmark” genes known to be genetically associated with human cancer. The identified SNPs often did not occur at the same position, as those described in human cancer due to variations between human and bovine genomes. All the SNPs found in cancer-related genes are noteworthy and they might impact the malignant phenotype of *Theileria*-infected leukocytes. What was particularly striking was non-synonymous SNPs in 12 of the COSMIC tier-1 group signature genes. These signature genes might define the genetic predisposition of *Theileria*-infected bovine leukocytes to develop a cancer-like phenotype, because of their established role as oncogenes (*FLT4*, *NOTCH2*, *MAP3K1*, *DAXX*, *FCGR2B*, *ROS1*), tumor suppressor gene (TSG) (*BARD1*, *KMT2C*, *GRIN2A*, *BAP1*), or fusion genes (*SLC34A2*, *NOTCH1*)³⁷. Among the 12 genes examined, homozygous mutations were discovered in *GRIN2A*, *MAP3K1*, *KMT2C*, and *DAXX*^{38,39}. TSG inactivation due to homozygous mutations or oncogene activation due to heterozygous mutations may contribute to the cancer-like features of *Theileria*-infected leukocytes. Loss of activity of *GRIN2A*, a well-known TSG, and activation of oncogenes (*MAP3K1*, *DAXX*, *FLT4*, *NOTCH2*, *FCGR2B*, *ROS1*) might contribute to *T. annulata*-induced transformation in infected leukocytes.

SNPs occurred in genes implicated in the genomic instability pathway (*DAXX*, *BARD1*, *KMT2C*, and *BAP1*), which is critical for cancer development, but has never been investigated in *Theileria*-induced leukocyte transformation. Genomic instability is a characteristic trait of nearly all cancer types and may be induced by endogenous (e.g., replication issues), or external (e.g., radiation) agents, and it's almost always the consequence of deficient or abnormal DNA repair processes⁴⁰. SNPs were observed in DNA repair and replication pathway genes that are often mutated in cancer, and their presence increases the likelihood of host genome instability in *Theileria*-infected leukocytes. One of the essential DNA repair genes, *CHEK2*, harbored SNPs in all six *Theileria*-infected leukocyte cell lines. *CHEK2* is a cell cycle checkpoint kinase that acts as a tumor suppressor, and defects in *CHEK2* have been linked to an increased risk of cancer⁴¹. *ROS1* had the highest number of SNPs in all six *Theileria*-infected leukocyte cell lines, so taking a hint from cancer research, we used crizotinib to inhibit the activity of bovine *ROS1*. Crizotinib inhibited the proliferation of *Theileria*-infected leukocytes that subsequently died. Overall, SNPs were identified in genes that should improve understanding of the cancer-like phenotype of *Theileria*-infected leukocytes and initiate the development of targeted therapies for the treatment of tropical theileriosis.

SNP signatures were extracted from the WGS datasets and compared to established cancer signatures to understand the basis of somatic mutation and genomic instability in *Theileria*-infected and transformed leukocytes. This uncovered two main cancer-like signatures indicative of a defective MMR and 5-methyl cytosine (5mC) deamination. MMR involves 5mC deamination-induced mismatches, and its absence in *Theileria*-infected leukocytes may result in an SBS1A mutational signature⁴². Mutations in MMR pathway genes and polymerase proofreading enzymes (DNA Polymerase Epsilon, Catalytic Subunit (*POLE*) and DNA polymerase delta 1, catalytic subunit (*POLD1*)) are all linked to defects in DNA replication³⁴. *POLE* and *POLD1* genes that play a critical role in DNA replication and repair harbored multiple SNPs in all six *Theileria*-infected leukocyte cell lines. In cancer cells mutations in DNA repair genes underlie many forms of genomic instability, including MSI and chromosomal instability (CIN)⁴³. MSI was observed in all *Theileria*-infected leukocyte lines indicating that parasite infection may generate host genomic instability, most likely due to defects in DNA mismatch repair activities. Mutations in the *DAXX* gene that can operate as either a TSG or an oncogene have been demonstrated to affect chromosomal stability and telomere maintenance in cancer^{35,44}. *DAXX* mutations were consistently detected, and loss of *DAXX* protein and alternative telomere extension may be related to CIN in *Theileria*-infected leukocytes. In cancer, negative regulation of *DAXX* dampens the cellular apoptotic response and is linked to Pin1-mediated prolyl isomerization. Pin1 is a conserved pathway in *Theileria*-infected leukocytes to regulate host oncogenic signaling^{45,46}.

We queried our dataset for CNVs and SVs, two forms of alteration that are known to play a significant role in cancer and influence a greater percentage of the genome than SNPs⁴⁷. Ten cancer-related genes had copy number alterations in all six *T. annulata*-infected leukocyte cell lines. Amplification of specific loci can impact the expression of *NFE2L2*, *BCL2L1*, *PAX8*, *PARP10*, *MDM4*, or *BPTF* to promote the stimulation of infected leukocyte proliferation^{48,49}. The activation of transcription factors NRF2 (*NFE2L2*), *PAX8* or *BPTF* may be just as crucial during the cellular transformation induced by *Theileria* infection, as the activation of other transcriptional regulators such as Myc (c-Myc), Nuclear Factor kappa B (NF- κ B), AP1 has been described^{5,50}. Differential expression of the NRF2, *PAX* and *PARP* gene families have been linked to *Theileria*-induced alterations in the infected host leukocyte⁵. We also noticed *NFE2L2* (*NRF2*) and *BCL2L1* overexpression in the six *Theileria*-infected leukocyte cell lines. Given their importance in cancer and their function as transcription factors and anti-apoptosis pathways, we hypothesize that these factors could potentially play a significant role in the immortalization of the bovine leukocytes. Furthermore, our study demonstrates that leukocytes infected by *T. annulata* exhibit aneuploidy-like cancer cells. This condition could potentially be triggered by copy number variations (CNV), or genome instability in infected leukocytes. Both aneuploidy and chromosomal instability have been widely acknowledged as fundamental characteristics of cancer⁵¹. We

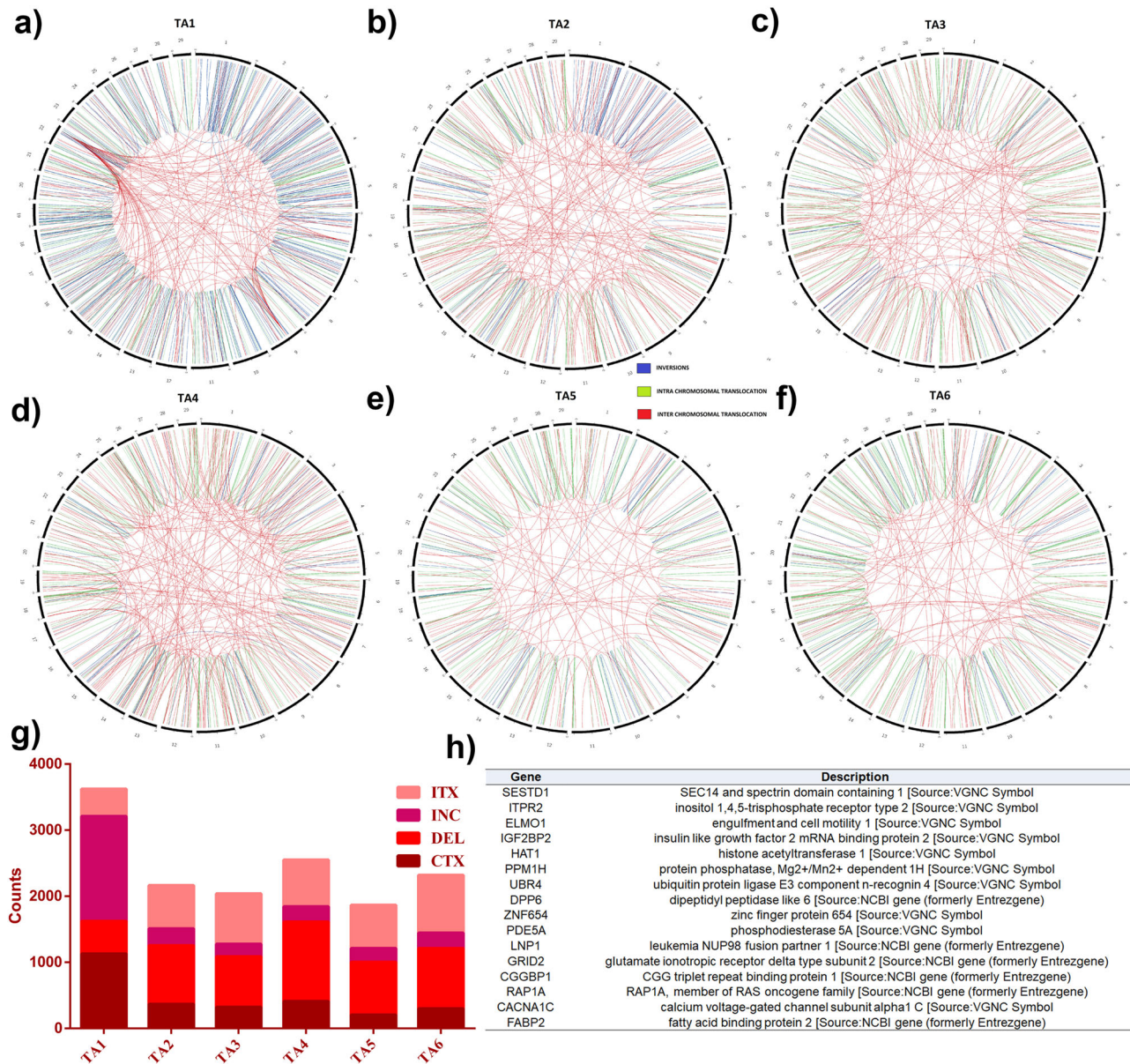


Fig. 7 | Distribution of structural variations across six *T. annulata* infected samples. a–f Circos plots display the structural variations (SVs) identified in individual samples (TA1–TA6). Chromosomes are arranged in a circular layout from 1 to 29. Lines within the plots represent different types of SVs: blue lines indicate inversions, green lines denote intra-chromosomal translocations, and red lines represent inter-chromosomal translocations. **g** The bar graph represents the number

of structural variations (SVs) across samples (TA1–TA6). Each bar is divided into four categories: inversions (INV), intra-chromosomal translocations (ITX), inter-chromosomal translocations (CTX), and deletions (DEL). Each color code represent different type of SVs. **h** Table presents a list of genes that exhibit common structural variations across all six samples.

hypothesize that copy number alterations are critical events in the evolution of the cancer phenotype in the *Theileria*-infected leukocytes, given the above observations of highly conserved CNVs across all six infected cell lines.

Theileria-infected leukocyte cell lines exhibit irreversible and reversible variations relating to changes in gene expression and epigenetic processes that contribute to developing a malignant phenotype^{1,50}. Like cancer, it is possible that genetic and epigenetic mechanisms are not independent events, but they interconnect and benefit each other to contribute to the cancer-like phenotypes of *Theileria*-infected leukocytes⁵². One of the bovine genes found to be altered, KMT2C, is known to control the activity of DNMT3A a de novo DNA methyltransferase and silencing of its expression via histone methylation has been associated with cancer metastasis⁵². Although we did not determine if the somatic alterations in key cancer-related genes affected epigenetic regulation, our results highlight the complexity of changes that could result from such alterations.

Research has indicated that more than half of human cancers have mutated TP53 genes in their somatic cells^{53,54}. However, our study did not reveal any mutations in this gene. This is in line with earlier work, which identified buparvaquone (BPQ) as a viable treatment for parasites, resulting in activation of TP53 and death of host leukocytes⁵⁵. Additionally, our discovery of Nutlin 3a, which activates the TP53 gene and terminates the infected leukocytes, is in agreement with this. While SNPs have been identified in genes regulating TP53's activity, such as TP53RK and TP53TG5, it remains to be seen if these SNPs affect TP53's ability to induce leukocyte apoptosis. We hypothesize that *Theileria* infection inactivates TP53 by binding it to Mdm2, thereby permitting the host cells to become immortalized⁵⁴. We believe it is the TP53 gene that plays a major role in immortalizing the host cells, not the parasite.

Pathogens such as *Epstein-Barr virus* (EBV), *hepatitis C virus*, *HIV*, *Helicobacter pylori*, and *Plasmodium falciparum* have been associated with a

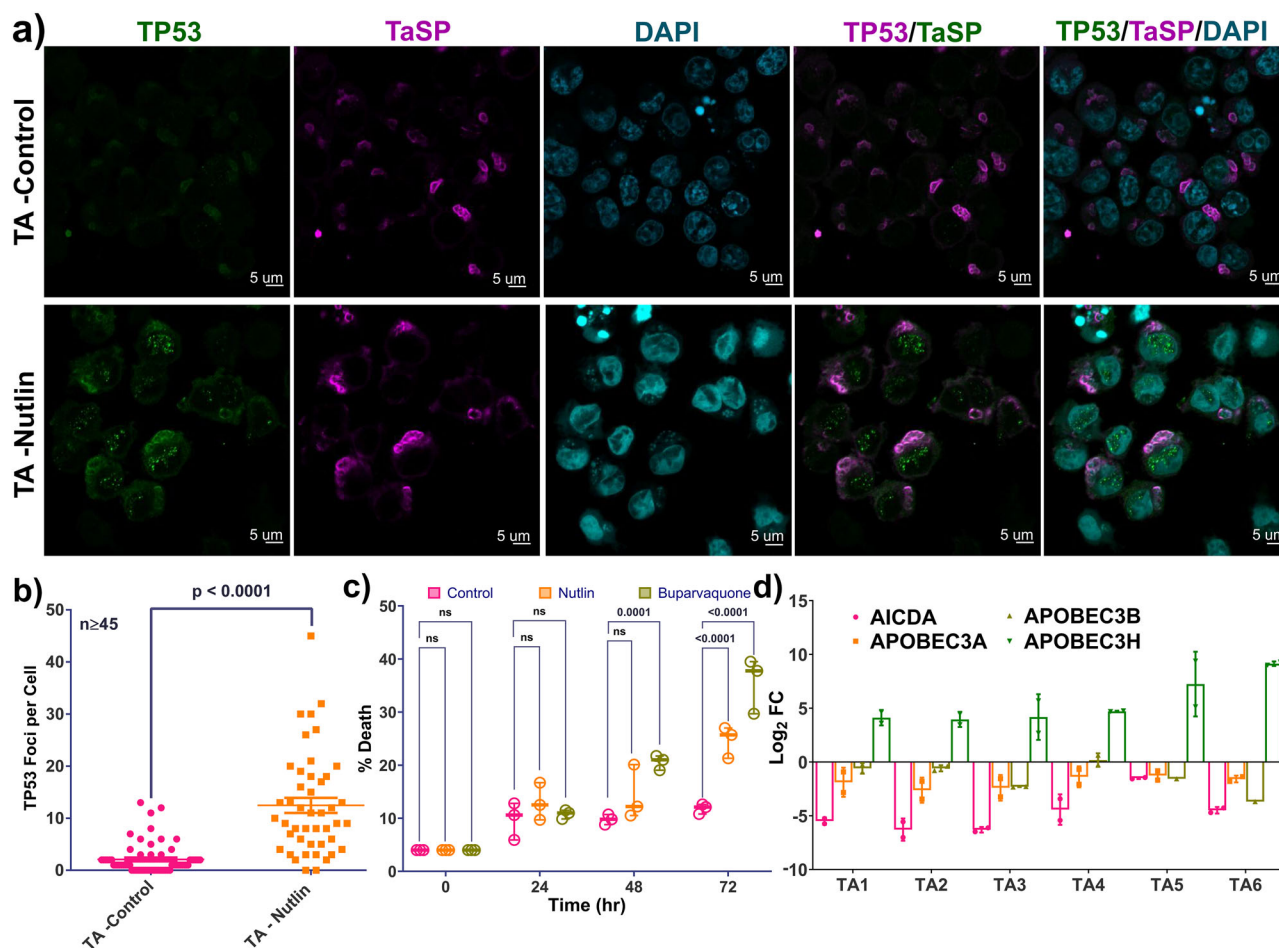


Fig. 8 | Effect of Nutlin-3A treatment on *T. annulata* infected bovine lymphocytes. **a** Immunofluorescence analysis of *T. annulata* infected cells after 48 hrs treatment with Nutlin-3A using anti-p53 (Green) and TaSP antibody (Magenta). DAPI was used for nuclear staining. **b** The graph demonstrates the activation of TP53 post Nutlin-3A treatment in *T. annulata* infected cells. **c** The box-plot depicts the percentage of dead *T. annulata* infected cells post Nutlin-3A treatment at 0 hr, 24 hr 48 hr and 72 hr. Error bars represent the mean with standard deviation (SD).

Dunnett's multiple comparison test was applied w.r.t the control group for the treated group at each time point, where p-value < 0.01 was considered to be significant. **d** Expression levels of AID family genes in *T. annulata* infected cells: qPCR-based quantification of AID family genes in *T. annulata* infected cells in comparison to uninfected healthy PBMCs. Error bars indicate the mean \pm standard deviation (SD), showing the variation across sample.

greater risk of B cell lymphoma and other cancers³⁶. Upregulated expression of the APOBEC3H gene suggests a resemblance to cancer, whereby this cytidine deaminase might contribute to the increased number of SNPs observed in *Theileria*-infected leukocytes. In addition, the APOBEC3 protein also plays a crucial role in the innate immune response to various viruses linked to the development of lymphomas, such as *Epstein-Barr virus* (EBV) and human T-cell leukemia virus (HTLV)⁵⁶. Several studies have demonstrated the critical role of APOBEC3H, APOBEC3B, and APOBEC3A in the formation of mutations in numerous cancer types, like head/neck, lung, cervical, and bladder cancers^{57–59}. However, further research is needed to understand better the connection and mechanism of the somatic mutations and cytidine deaminase related to *T. annulata*-induced bovine leukocyte transformation.

This study provides the first comprehensive genomic characterization of *T. annulata*-transformed bovine leukocytes, revealing critical insights into parasite-driven oncogenesis while identifying important limitations that guide future research directions. The absence of matched healthy controls from the same animals represents a key methodological constraint, as our reliance on the *Bos taurus* reference genome cannot account for individual-specific structural variations, pre-existing copy number alterations, or host genetic polymorphisms that may influence transformation susceptibility. Future studies incorporating paired samples would enable more precise discrimination between *Theileria*-induced somatic alterations

and innate host genomic variability. Another limitation of our study is that we compared SNP distribution in the genomes of 6 constantly proliferating cultured cell lines, and as a consequence, any SNP that negatively impacted on infected leukocyte proliferation would not be seen, because such cells would drop out of the continuous in vitro culture. It argues that the SNPs we identified don't affect proliferation and a case in point is ROS1 discussed below.

Even though ROS1 harbored the highest number of SNPs ROS1 displayed was significant upregulated expression in *Theileria*-infected leukocytes compared to non-infected PBMC controls. Its role in infected leukocyte proliferation was validated through pharmacological inhibition of its tyrosine kinase activity with crizotinib. A critical next step would be CRISPR-based knockout/knockdown of ROS1, coupled with pathway rescue experiments to establish definitively that ROS1 plays an essential role in *Theileria*-induced leukocyte proliferation. Our use of cultured cells additionally limits direct extrapolation to natural infection scenarios, highlighting the need for complementary in vivo validation in experimentally infected cattle.

Our whole-genome analysis uncovered widespread somatic mutagenesis and infection-associated genomic instability, likely mediated by APOBEC3H overexpression and subsequent mutations in DNA repair genes. While some genes (e.g., *ROS1*) accumulated numerous SNPs, others (e.g., TP53) remained unaffected—a phenomenon requiring further investigation.

The parasite's essential role in transformation was demonstrated by rapid leukocyte death following BPQ- or nutlin-3a-induced TP53 reactivation, supporting our hypothesis that TP53 sequestration permits excessive SNP accumulation. This TP53-dependent cell death may result from lethal mutational burden. Notably, ROS1 emerged as a potential therapeutic target, and DNA damage repair pathways may offer new treatment opportunities. Our genome-wide analysis of six independent *Theileria*-transformed leukocyte lines provides critical insights into protozoan-driven oncogenesis and establishes a foundation for host-directed therapies to augment current antiparasitic drugs. Future work must translate these in vitro findings into clinical applications while deciphering the complex host-parasite interactions underlying malignant transformation.

Materials and methods

DNA samples

The blood samples (<5 mL) were collected from cross breed cattle (*Bos taurus*, female and 4–5 years old) with confirmed *T. annulata* infections from different states of India. We confirm that the biological material was obtained exclusively under a non-research context, as the whole blood samples were collected by a veterinary officer for standard diagnostic and health monitoring procedures; the PBMCs were subsequently utilized as excess material from these routine clinical actions. Five distinct transformed leukocyte cultures were successfully established from these clinical samples, all demonstrating stable growth patterns and typical transformation markers. The experimental design also included the Rakshavac-T reference cell line; a commercially available attenuated *T. annulata* strain widely used in Indian bovine vaccination initiatives. All cell cultures ($n = 6$) were grown in their specific optimized media until achieving sufficient cell numbers for molecular studies⁶⁰. During the 6–8-week cultivation period (15–30 passages), we systematically monitored three key parameters: Consistent logarithmic-phase proliferation rates, maintenance of intracellular parasite loads and development of characteristic transformed cellular morphology.

The DNA was extracted using QIAamp DNA Mini Kit (Qiagen) from the in vitro *T. annulata*-infected bovine cells, according to the manufacturer's recommendations. Only isolated DNA with A260/280 ratios above 1.8 and proven high quality by gel electrophoresis were used for sequencing. All six samples were used for the WGS sequencing. Following the quality check, DNA samples were kept at -80°C until further studies. Polymerase chain reaction (PCR) utilizing primers specific for the *Theileria annulata* surface protein (TaSP) gene of *T. annulata* indicated the presence of parasites in host cells. The following are the PCR conditions for the aforementioned primers:

TaSP: 95°C for 3 min, followed by 35 cycles of 95°C for 1 min, 55°C for 1 min, 72°C and 1 min and a final extension of 5 min at 72°C 5 min.

WGS analysis of *T. annulata*-infected bovine cells

The DNA from *T. annulata*-infected bovine cell lines TA1, TA2, TA3, TA4, TA5, and TA6 were submitted for sequencing with at least 30x coverage on the Illumina HiSeq sequencer (150×2 library type) using standard methods. We trimmed raw reads for adapter sequence removal, and reads with a minimum read length of 50 bp, and minimum base quality Q30 were selected using the Trimgalore-v0.4.4 tool. The adapter free good quality reads were mapped to reference *Bos taurus* genome (Ensembl database: ARS-UCD1.2:CM008168.2) using Bowtie2-v2.0.5 at default parameters to generate alignment data (BAM format). The alignment data was processed using the Samtools-v1.9 program to generate mpileup data. We used Varscan2 (mpileup2snp and mpileup2indel) to predict variants (SNPs/indels) with various parameters such as minimum coverage (read depth) of 10, varfreq (variant frequency) of 0.2, Phred base score of 30 (average quality score), and p-value threshold of 0.01 after filtering out insignificant and PCR duplicate reads⁶¹. The predicted variants (SNPs and INDELs) were annotated based on the reference genome and gene feature information using the SnpEff-v3.3 h tool. The predicted SNPs were reported as homozygous (1/1) if the variant allele frequency was a minimum of 60%. SnpEff and Ensembl Variant Effect Predictor⁶² were used to determine the effect of these

variations on the genes (VEP). SnpEff annotates genomic variants and coding effects such as synonymous or non-synonymous amino acid substitution, start or stop codon gains or losses, or frameshifts⁶³. VEP employs the Sorting Intolerant From Tolerant (SIFT) method to forecast the impact of variations on the gene. SIFT spans from 0 to 1 and is based on a normalized chance of spotting the new amino acid at that place. A number between 0 and 0.05 is likely to impact protein function negatively. The WGS data generated in this study is submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA914920.

COSMIC and TCGA cancer database

We utilized two primary human cancer databases - COSMIC (Catalog of Somatic Mutations in Cancer) and TCGA Pan-Cancer Atlas (<http://tcgportal.org/>) - for comparative genomic analysis of *Theileria*-transformed bovine leukocytes, given the 85–90% conservation of cancer pathways between species and absence of comprehensive bovine-specific resources⁶⁴. Whole genome sequencing variants were cross-referenced with these databases to identify conserved cancer-associated mutations and pathway enrichments (Reactome analysis with bovine ortholog mapping). All candidate genes underwent bovine-specific validation through qPCR expression profiling, pharmacological inhibition assays, and flow cytometric analysis of proliferation/apoptosis. This cross-species approach was justified by documented functional conservation of core oncogenes (TP53, ROS1) and cancer hallmarks (NF- κ B activation, MMP9 upregulation, TP53 suppression) observed in our model¹. Human data served solely for initial target prioritization, with all mechanistic conclusions derived from bovine experimental validation. We acknowledge limitations including potential undiscovered species-specific differences, selection bias toward proliferation-enhancing mutations in cultured cells, and gaps in bovine genome annotation. This strategy enabled systematic investigation of *Theileria*-induced transformation mechanisms despite limited species-specific resources, while establishing a framework for future studies as bovine omics databases develop.

Mutational signature and microsatellite instability

To find out the prevalence of any mutation type in the *T. annulata* infected samples we converted all mutations from the WGS data sets into a matrix composed of 96 single base substitutions for each mutation type (C > A, C > G, C > T, T > A, T > C and T > G) using each possible 5' and 3' contexts for all samples by applying the R package (SomaticSignatures)⁶⁵. The distribution of mutational proportion for every 96 types are provided in the Supplementary Data 2. After identifying the proportion of 96 mutational types we applied the R package deconstructSigs to infer the actual mutational signature⁶⁶. In total, 30 signatures reported by COSMIC are included in the analysis. To further validate the signature of mismatch repair deficient, we ran MSIsensor2⁶⁷ to identify the MSI level across *T. annulata* infected cells⁶⁷. As suggested by the literature, MSI scores greater than 20 were regarded as MSI, and values less than 20 were considered microsatellite stable.

Copy number variation and structural variation

CNVs, or copy number variants, were identified using CNVcaller⁶⁸. We used 800 bp overlapping sliding windows to achieve this, and the results were compared across the samples. A population read depth file is created from the averaged read depths of all samples, and from this file, the potential CNV windows are selected. The potential CNV windows are merged into the CNV area if the distance between the two initial calls is less than 20% of their total length and the Pearson's correlation index of the two CNVRs is significant at the $0 = 0.01$ level (CNVR). Sample clustered read depth and individual integer copy numbers were used to identify the variant genotype. Two statistical analysis Silhouette score and the Calinski-Harabasz score, were used to find the significant CNVs regions. Structural variants were computed using Breakdancer to call for large inter and intrachromosomal changes⁶⁹. The Breakdancer variant caller identified longer variants (>1KB), such as inversion, duplication, translocation, insertion, and deletion. Only

Table 1 | List of Primers and antibodies used in the study

PRIMER SEQUENCE (5' - > 3')		
GENES	FORWARD PRIMERS	REVERSE PRIMERS
<i>GRIN2A</i>	CGTGCCGGAAGTTTGTCAAG	CCGTTCCACACGTTGTTCAC
<i>MAP3K1</i>	CGGCCAGTTGTAGACACCAT	CTGCTTGGCCTTTGCACAAAT
<i>NOTCH1</i>	CTATCTCTGCTTCTGCCTCAAG	GTAGCCGCTCAATCTTGTCCAG
<i>NOTCH2</i>	AGGAACCTGCCTTGATGACG	AGCACACAGGCAGGAGTAAC
<i>KMT2C</i>	CCCATCCTTCGAGGGCTTTT	AGGCTGGGAATATGGGTCCT
<i>DAXX</i>	CCCCTCCAGTGTAGTTGCTG	GGGGCAAGGCTTCAATCTCT
<i>BARD1</i>	AGGGCCAGAGTCTTCAGGAG	GCAATCTCAGCACTCAGCAC
<i>BAP1</i>	ATCGGCAGCGCTTTCAATTC	GGCTGTCGCTCTCTCCAAAA
<i>SLC34A2</i>	ATTCTCTGCGTCTTCCAAGG	ATTCCCAGCTACCTTTCTCTC
<i>AICDA</i>	ACTTTCAAAGCCTGGGAGGG	GCATCCCGCAAGTCATCAAC
<i>APOBEC3A</i>	AAGCACTGCTGGGAAACCTT	CTTTAGAGAGGCTGGCGTCC
<i>APOBEC3B</i>	GTCTATGGTGGGACTCGCTG	AGGTGATGTGGTAACGCTCG
<i>APOBEC3H</i>	ACCAGCTGAAGCAGCGTAAT	TTTGTAGCTCTGGCTCGGG
<i>BCL2L1</i>	AATGAACTCTTCCGGGACGG	TTGTCTACGCTTTCCACGCA
<i>NFE2L2</i>	CTACGGGCAAAAGCTCTCCA	TCTGCAATTCTGAGCAGCCA
Antibody Details		
	Catalog No.	Intender
ROS1 (D4D6 ⁺) Rabbit mAb	3287	Cell Signaling Technology
p53 Mouse Monoclonal Antibody (DO-7)	MA5-12557	Invitrogen
TaSP Rabbit pAb	Commercially synthesized from Genscript	GenScript

structural variations with a confidence score of 99 were retained to maintain SVs with high confidence.

Real-time PCR

We performed Real-time PCR (q-PCR) by isolating RNA from *Theileria*-infected cells using the MN RNA isolation kit according to the manufacturer's instructions. cDNA was synthesized from the RNA using reverse transcriptase (Clontech). cDNA isolated from the PBMCs of the healthy animal was used as a control for the q-PCR. For q-PCR, primers for the following ten genes were designed and synthesized. The sequences of the primers are listed in Table 1. Using the BioRad machine's default settings, we examined the relative expression of the target genes using HPRT as the reference gene. $2^{-\Delta\Delta CT}$ was calculated to evaluate the gene's expression by comparing cDNA from healthy animals and *Theileria*-infected cells. Each experiment was conducted in triplicate.

Cell culture and drug susceptibility assay

T. annulata infected cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum at 37 °C with 5% CO₂⁶⁰. A challenge experiment with Crizotinib was carried out to assess the susceptibility of *Theileria*-infected cells to drug pressure (Cat no:12087; Cayman Chemicals). After 48 h, each well received 20 µL of 1.5 mM resazurin dye, and the fluorescence intensity of the cells at 570 nm was measured 8 h later to test cell survival. Untreated cells were used as a control. Likewise, host cell cytotoxicity experiments were performed on PBMCs collected from healthy cattle that tested negative for *T. annulata* infection. Each experiment was repeated three times.

Flow cytometry

To investigate potential aneuploidy in *T. annulata*-infected leukocytes, we employed a standardized DNA flow cytometry protocol comparing two infected cell lines with healthy bovine peripheral blood mononuclear cells (hPBMCs) as diploid controls. hPBMCs were isolated from healthy cattle blood and fixed in 70% ice-cold ethanol for preservation. The infected cell lines underwent identical fixation procedures to ensure methodological

consistency. Prior to analysis, all samples were washed twice with DPBS to remove residual ethanol and stained with 300 µl propidium iodide solution (50 µg/ml PI, 0.1% Triton X-100) for 30 min under dark conditions to prevent fluorophore degradation.

Flow cytometric measurements were conducted using a BD cytometer with instrument parameters carefully calibrated using hPBMC controls. Voltage settings for forward scatter (FSC) and side scatter (SSC) were optimized to capture the complete cell population distribution, while PI fluorescence detection parameters were adjusted for precise DNA content quantification. Each sample analysis captured 50,000 cellular events, with minor FSC/SSC adjustments made for infected cells to account for their distinct morphological characteristics. FlowJo software facilitated comprehensive data analysis through a multi-step analytical pipeline. Initial quality control measures included doublet exclusion via side scatter area versus height (SSC-A vs SSC-H) gating and lymphocyte population isolation through FSC-A vs SSC-A parameters. DNA content histograms were generated from PI fluorescence intensity distributions, with cell cycle phase determination performed using FlowJo's integrated Watson (Pragmatic) algorithm. The DNA index (DI) was calculated as the ratio of G1 peak mean fluorescence intensities between infected cells and diploid hPBMC controls, with values between 1.1 and 1.9 considered indicative of aneuploidy according to established criteria²⁸. The gating criteria has been mentioned in Supplementary Fig. 3

IFA

IFA experiments were done to verify the presence of *T. annulata* parasites and TP53 expression in the B-lymphocytes cells using a TASP antibody (1:200) raised against the *Theileria annulata*-surface protein and TP53 antibody (1:100)⁷⁰. The IFA experiment and analysis were conducted following the methods described in a previous study⁷⁰. In summary, 5×10^5 *T. annulata* infected cells were cultured and subsequently pelleted. The pelleted cells were washed three times with 1X PBS and then fixed with 4% paraformaldehyde at 37 °C for 10 min. Following fixation and permeabilization (0.1% Triton X-100), cells were then incubated in a blocking buffer (2% BSA in 1X PBS) at room temperature for 1 h. Subsequently, the cells

were exposed to primary antibodies overnight at 4 °C. After the washing step; the cells were incubated with a secondary antibody at room temperature for 1 h. Following another round of washing, the DNA within the cells was labeled with DAPI. Finally, the samples were mounted using mounting media and images were captured using a fluorescent microscope. The acquired images were processed and quantified using the ZEN 3.3 (blue edition) software program.

Statistics and reproducibility

In this study, the statistical analysis was done using GraphPad Prism 10 software (GraphPad, CA, USA). For comparisons between two groups such as healthy PBMCs and *T. annulata*-infected cells, or control versus treated groups, unpaired two-way ANOVA was employed. To assess statistical significance between BPQ- or Nutlin-treated cells and the control (untreated) group at various time points, Dunnett's multiple comparison test was applied. For mutation profile analysis, one-way ANOVA was conducted to compare the distribution of six types of single-base substitutions (SBSs) across all samples. Tukey's multiple comparisons test was used post hoc to determine statistical significance among the SBS types.

To ensure data reproducibility all experiments were performed in biological replicates and in multiple samples.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The WGS data that support the findings of this study have been deposited in NCBI with the accession code [PRJNA914920](https://www.ncbi.nlm.nih.gov/submit/PRJNA914920). The remaining numerical source for the graphs has been provided in Supplementary_Data 6.xls. Full blots of the figure panel have been shown in Supplementary Fig. 4.

Received: 20 November 2024; Accepted: 6 October 2025;

Published online: 20 November 2025

References

1. Tretina, K., Gotia, H. T., Mann, D. J. & Silva, J. C. Theileria-transformed bovine leukocytes have cancer hallmarks. *Trends Parasitol.* **31**, 306–314 (2015).
2. Li, Z. et al. Theileria annulata subtelomere-encoded variable secreted protein-TA05575 binds to bovine RBMX2. *Front Cell Infect. Microbiol.* **11** <https://www.frontiersin.org/articles/10.3389/fcimb.2021.644983> (2021).
3. Villares, M. et al. Trifloxystrobin blocks the growth of Theileria parasites and is a promising drug to treat Buparvaquone resistance. *Commun. Biol.* **5**, 1–11 (2022).
4. Cheeseman, K. et al. Dynamic methylation of histone H3K18 in differentiating Theileria parasites. *Nat. Commun.* **12**, 3221 (2021).
5. Kinnaird, J. H. et al. A Bovine lymphosarcoma cell line infected with theileria annulata exhibits an irreversible reconfiguration of host cell gene expression. *PLoS One* **8**, e66833 (2013).
6. Woods, K., Perry, C., Brühlmann, F. & Olias, P. Theileria's strategies and effector mechanisms for host cell transformation: from invasion to immortalization. *Front Cell Dev. Biol.* **9**, 662805 (2021).
7. Weitzman, M. D. & Weitzman, J. B. What's the damage? The impact of pathogens on pathways that maintain host genome integrity. *Cell Host Microb.* **15**, 283–294 (2014).
8. Yao, Y. & Dai, W. Genomic instability and cancer. *J. Carcinog. Mutagen.* **5**, 1000165 (2014).
9. Huang, R. & Zhou, P.-K. DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduct Target Ther.* **6**, 1–35 (2021).
10. Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16**, 168–174 (2000).
11. Certad, G. et al. Fulminant cryptosporidiosis associated with digestive adenocarcinoma in SCID mice infected with *Cryptosporidium parvum* TUM1 strain. *Int. J. Parasitol.* **40**, 1469–1475 (2010).
12. Watson, I. R., Takahashi, K., Futreal, P. A. & Chin, L. Emerging patterns of somatic mutations in cancer. *Nat. Rev. Genet.* **14**, 703–718 (2013).
13. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330–1334 (2017).
14. Roy, S. et al. Population genetic analysis of the *Theileria annulata* parasites identified limited diversity and multiplicity of infection in the vaccine from India. *Front Microbiol.* **11**, 579929 (2020).
15. Sharma, Y. et al. A pan-cancer analysis of synonymous mutations. *Nat. Commun.* **10**, 2569 (2019).
16. Shen, X., Song, S., Li, C. & Zhang, J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature.* **606**, 725–731 (2022).
17. Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
18. Abbas, T., Keaton, M. A. & Dutta, A. Genomic instability in cancer. *Cold Spring Harb. Perspect Biol.* **5**, a012914 (2013).
19. Plass, C. et al. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat. Rev. Genet.* **14**, 765–780 (2013).
20. Shaw, A. T. et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N. Engl. J. Med.* **371**, 1963–1971 (2014).
21. Degasperi, A. et al. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* (1979) **376**, abt9283 (2022).
22. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
23. Islam, S. M. A. et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* **2**, 100179 (2022).
24. Aleksandrov, R. et al. Protein dynamics in complex DNA lesions. *Mol. Cell* **69**, 1046–1061.e5 (2018).
25. Pečina-Šlaus, N., Kafka, A., Salamon, I., Bukovac, A. Mismatch repair pathway, genome stability and cancer. *Front. Mol. Biosci.* **7** Available from: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.00122> (2020).
26. Shao, X. et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med. Genet.* **20**, 175 (2019).
27. Ippolito, M. R. et al. Gene copy-number changes and chromosomal instability induced by aneuploidy confer resistance to chemotherapy. *Dev. Cell* **56**, 2440–2454.e6 (2021).
28. Risques, R. A. et al. Redefining the Significance of Aneuploidy in the Prognostic Assessment of Colorectal Cancer. *Lab Invest* **81**, 307–315 (2001).
29. Li, L.-Y., Guan, Y., Chen, X.-S., Yang, J.-M. & Cheng, Y. DNA repair pathways in cancer therapy and resistance. *Front Pharmacol.* **11**, 629266 (2020).
30. Zhang, Y.-L., Wang, R.-C., Cheng, K., Ring, B. Z. & Su, L. Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer Biol. Med.* **14**, 90–99 (2017).
31. Haller, D. et al. Cytoplasmic sequestration of p53 promotes survival in leukocytes transformed by {Theileria}. *Oncogene* **29**, 3079–3086 (2010).
32. Hayashida, K. et al. MDM2 regulates a novel form of incomplete neoplastic transformation of *Theileria parva* infected lymphocytes. *Exp. Mol. Pathol.* **94**, 228–238 (2013).
33. Biegging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14**, 359–370 (2014).

34. Haradhvala, N. J. et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.* **9**, 1746 (2018).
35. Marinoni, I. et al. Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology* **146**, 453–460.e5 (2014).
36. Robbiani, D. F. et al. Plasmodium infection promotes genomic instability and AID-dependent B cell lymphoma. *Cell* **162**, 727–737 (2015).
37. Bamford, S. et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
38. Varmus, H. E. The molecular genetics of cellular oncogenes. *Annu. Rev. Genet.* **18**, 553–612 (1984).
39. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
40. Yoshioka, K.-I., Kusumoto-Matsuo, R., Matsuno, Y. & Ishiai, M. Genomic instability and cancer risk associated with erroneous DNA repair. *Int. J. Mol. Sci.* **22**, 12254 (2021).
41. Stolz, A., Ertych, N. & Bastians, H. Tumor suppressor CHK2: regulator of DNA damage response and mediator of chromosomal stability. *Clin. Cancer Res.* **17**, 401–405 (2011).
42. Fang, H. et al. Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. *Sci. Adv.* **7**, eabg4398 (2021).
43. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
44. Heaphy, C. M. et al. Altered telomeres in tumors with ATRX and DAXX mutations. *Science* (1979) **333**, 425 (2011).
45. Ryo, A. et al. A suppressive role of the prolyl isomerase Pin1 in cellular apoptosis mediated by the death-associated protein Daxx *. *J. Biol. Chem.* **282**, 36671–36681 (2007).
46. Marsolier, J. et al. Theileria parasites secrete a prolyl isomerase to maintain host leukocyte transformation. *Nature* **520**, 378–382 (2015).
47. Belzen IAEM van, Schönhuth, A., Kemmeren, P. & Hehir-Kwa, J. Y. Structural variant detection in cancer genomes: computational challenges and perspectives for precision oncology. *NPJ Precis. Oncol.* **5**, 1–11 (2021).
48. Yin, J. et al. Copy-number variation of MCL1 predicts overall survival of non-small-cell lung cancer in a Southern Chinese population. *Cancer Med.* **5**, 2171–2179 (2016).
49. Richart, L. et al. BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis. *Nat. Commun.* **7**, 10153 (2016).
50. Chaussepied, M. & Langsley, G. Theileria transformation of bovine leukocytes: a parasite model for the study of lymphoproliferation. *Res. Immunol.* **147**, 127–138 (1996).
51. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
52. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin?. *Cancer Cell* **22**, 9–20 (2012).
53. Zhu, G. et al. Mutant p53 in cancer progression and targeted therapies. *Front. Oncol.* **10**. Available from: <https://www.frontiersin.org/articles/10.3389/fonc.2020.595187> (2020).
54. Muller, P. A. J. & Vousden, K. H. p53 mutations in cancer. *Nat. Cell Biol.* **15**, 2–8 (2013).
55. Haller, D. et al. Cytoplasmic sequestration of p53 promotes survival in leukocytes transformed by Theileria. *Oncogene* **29**, 3079–3086 (2010).
56. Wagener, R. et al. Analysis of mutational signatures in exomes from B-cell lymphoma cell lines suggest APOBEC3 family members to be involved in the pathogenesis of primary effusion lymphoma. *Leukemia* **29**, 1612–1615 (2015).
57. Starrett, G. J. et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat. Commun.* **7**, 12918 (2016).
58. Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **45**, 977–983 (2013).
59. Burns, M. B. et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
60. Roy, S., Bhandari, V., Dandasena, D., Murthy, S. & Sharma, P. Genetic profiling reveals high allelic diversity, heterozygosity and antigenic diversity in the clinical isolates of the Theileria annulata from India. *Front. Physiol.* **10**, 673 (2019).
61. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
62. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
63. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
64. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
65. Gehring, J. S., Fischer, B., Lawrence, M. & Huber, W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**, 3673–3675 (2015).
66. Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium ICGC MML-Seq Consortium, et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
67. Niu, B. et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016 (2014).
68. Wang, X. et al. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience* **6**, 1–12 (2017).
69. Chen, K. et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods.* **6**, 677–681 (2009).
70. Barman, M. et al. Artemisinin derivatives induce oxidative stress leading to DNA damage and caspase-mediated apoptosis in Theileria annulata-transformed cells. *Cell Commun. Signal.* **21**, 78 (2023).

Acknowledgements

The authors gratefully acknowledge the Director of the National Institute of Animal Biotechnology (NIAB), Hyderabad, for providing essential resources and unwavering support throughout this study. We extend our sincere thanks to the Department of Biotechnology (DBT), Council of Scientific and Industrial Research (CSIR), and University Grants Commission (UGC) for awarding Junior Research Fellowships to D.D., A.S., S.S., V.M.A. and M.S.; D.D. also acknowledges the Regional Center for Biotechnology (RCB), Faridabad, for permitting him to pursue his Ph.D. research. This work was supported by the Department of Biotechnology (Grant No. BTPR11979/AAQ/1/608/2014) and the core funding of NIAB, Hyderabad. We also acknowledge the use of data generated by The Cancer Genome Atlas (TCGA) Research Network (<https://www.cancer.gov/tcga>).

Author contributions

P.S. and D.D. devised the study's idea, planned the experiments, and evaluated and interpreted the results. Most of the experiments were carried out by D.D. Data analysis was done by D.D., R.B., J.G., S.R., A.S., V.M.A., S.S., M.S., A.T., V.B. and P.S. The manuscript was written by D.D., A.S., V.M.A., S.S., M.S., V.B., and P.S. All authors were given a chance to discuss the findings and provide feedback on the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-09005-8>.

Correspondence and requests for materials should be addressed to Paresh Sharma.

Peer review information *Communications Biology* thanks Gordon Langsley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Nishith Gupta and Johannes Stortz.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025