RESEARCH ARTICLE

Paired RNA-Seq Dataset Revealing Differential Gene Expression in Canine Transmissible Venereal Tumour: A Pilot Study

Yaman S. Panchal[1]*, Abhi S. Patel[1], Bhomika A. Joshi[2], Ashish C. Patel[1], Umed V. Ramani[2], Subhash J. Jakhesara[2], Prakash G. Koringa[2]

Abstract

Canine transmissible venereal tumour (CTVT) is a contagious cancer that uniquely regresses spontaneously following vincristine chemotherapy. In this study, we analyzed a publicly available paired RNA sequencing dataset (SRA: ERR2044811-ERR2044822) comprising tumour biopsies collected from CTVT-afflicted dogs at day 0 (pre-treatment) and day 28 (post-treatment). Raw reads were processed with fastp for quality trimming and aligned to the canine genome (CanFam3.1) with HISAT2, followed by gene counting with HTSeq and differential expression analysis via the DESeq2 pipeline. Principal Component Analysis (PCA) separated day 28 (regressing) samples from day 0 (progressive) samples. We identified hundreds of differentially expressed genes (DEGs) (adjusted p-value <0.05) between day 28 and day 0. Notably, immune-related chemokine genes such as *CCL5* and *CXCL10* were strongly upregulated at day 28, whereas cell proliferation markers *MKI67* and *TOP2A* were markedly downregulated in regressed tumours. These expression changes highlight an activated immune response and reduced tumour cell proliferation during regression. Our findings corroborate the critical role of host immunity in CTVT regression and identify gene expression shifts accompanying vincristine-induced tumour remission.

Key words: Canine TVT, DEGs, Immune response, RNA-Seq, Tumour regression.

Ind J Vet Sci and Biotech (2026): 10.48165/ijvsbt.22.2.13

Introduction

Canine transmissible venereal tumour (CTVT) is one of the few known naturally transmissible tumours, affecting dogs worldwide (Murchison *et al.,* 2014). CTVT is spread by the physical transplantation of viable tumour cells during mucosal contact and typically manifests as genital or mucosal tumours (Belov *et al.,* 2015). Uniquely, this clonally derived tumour lineage can undergo spontaneous immune-mediated regression after a period of growth or it can be effectively cured by a short course of vincristine chemotherapy (Frampton *et al.,* 2018). Prior studies have linked CTVT regression with activation of T-cell mediated and cytokine-driven immune responses, along with induction of apoptosis and tumour suppressive mechanisms. However, the precise molecular drivers of regression remain only partially understood (Chiang *et al*., 2013). Recent genomic studies have provided insights into CTVT's origin and evolution and into its regression mechanisms. In particular, Frampton *et al.* (2018) profiled CTVT tumours before and after vincristine treatment and found that regression occurs in sequential steps involving innate immune activation, tissue remodelling, immune cell infiltration, and cell-cycle arrest.

In this paper, we present an analysis of a paired RNA-Seq dataset of CTVT, focusing on differential gene expression between pre-therapy (progressive phase) and post-therapy (regressing) tumour biopsies. We especially focus on key genes such as *CCL5, CXCL10, MKI67,* and *TOP2A*, which emerged as significant in our analysis and have known roles in immune response or cell proliferation. Our results shed further light on the molecular signature of regressing CTVT and reinforce the role of the host immune response in spontaneous tumour control.

[1]Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand-388001, Gujarat, India

[2]Department of Veterinary Biotechnology, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand-388001, Gujarat, India

Corresponding Author: Dr. Y. S. Panchal, Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand-388001, Gujarat, India. e-mail: dryamanpanchal@gmail.com

How to cite this article: Panchal, Y. S., Patel, A. S., Joshi, B. A., Patel, A. C., Ramani, U. V., Jakhesara, S. J., & Koringa, P. G. (2026). Paired RNA-Seq Dataset Revealing Differential Gene Expression in Canine Transmissible Venereal Tumor: A Pilot Study. Ind J Vet Sci and Biotech, 22(2), 71-76.

Source of support: Nil

Conflict of interest: None

Submitted 15/12/2025 **Accepted** 02/01/2026 **Published** 10/03/2026

Materials and Methods

Dataset Source

This study analysed a public RNA-Seq dataset of CTVT (ENA Project-PRJEB21960). The dataset consists of 12 libraries (ERR2044811-ERR2044822) derived from tumour biopsies of six dogs with CTVT. Each dog contributed two samples: one collected at Day 0, before treatment (progressive phase) and one at Day 28, after vincristine treatment (regressive phase). The raw sequencing data (paired-end FASTQ files) were downloaded from the NCBI Sequence Read Archive (SRA) under the accession ERP024276 (Table 1). We note that the data analyzed here were generated by a previous research group (Frampton *et al.,* 2018) and made publicly available. We used the data solely for secondary analysis. No new animal experiments were performed by us.

Table 1: Sample metadata for the paired RNA-Seq dataset from six CTVT cases

Dog ID	Time point	BioSample ID	SRA Run Accession
761	Day 0 (Tumour)	SAMEA104176492	ERR2044811
761	Day 28 (Post)	SAMEA104176492	ERR2044812
765	Day 0 (Tumour)	SAMEA104176493	ERR2044813
765	Day 28 (Post)	SAMEA104176493	ERR2044814
766	Day 0 (Tumour)	SAMEA104176494	ERR2044815
766	Day 28 (Post)	SAMEA104176494	ERR2044816
772	Day 0 (Tumour)	SAMEA104176495	ERR2044817
772	Day 28 (Post)	SAMEA104176495	ERR2044818
774	Day 0 (Tumour)	SAMEA104176496	ERR2044819
774	Day 28 (Post)	SAMEA104176496	ERR2044820
775	Day 0 (Tumour)	SAMEA104176497	ERR2044821
775	Day 28 (Post)	SAMEA104176497	ERR2044822

Note: Each CTVT case (Dog ID) has paired tumour biopsies collected at Day 0 (pre-treatment) and Day 28 (post-treatment/vincristine). BioSample and run accession IDs are provided for public retrieval of the sequencing reads.

Samples, RNA Extraction and Sequencing

CTVT tumour biopsy samples originated from naturally occurring cases in pet dogs (mixed breeds) treated at a veterinary hospital in Brazil (Frampton *et al.,* 2018). Six individual dogs (identified as CTVT 761, 765, 766, 772, 774 and 775 in the original study) were included. For each dog, a tumour biopsy was taken prior to therapy (day 0, representing an actively growing tumour) and a follow-up biopsy was taken at day 28, after the dog received three weekly doses of vincristine and the tumour had entered remission.

Total RNA was extracted from each biopsy using column-based RNA purification (Qiagen RNeasy Kit). mRNA was enriched via poly-A selection, and strand-non-specific cDNA libraries were constructed using Illumina TruSeq protocols with insert sizes of approximately 200 bp. Paired-end sequencing (2 × 75 bp) was performed on an Illumina NextSeq 550 platform for all libraries. Base calling and demultiplexing were performed with Illumina's software, producing raw FASTQ files for each sample's read 1 and read 2. All raw data used in this analysis were publicly available and not generated in-house.

Quality Control

Initial quality assessment of the raw reads was carried out using FastQC (v0.12.0) to evaluate read quality scores, GC content, length distribution, adapter content and duplication levels.

Read trimming and filtering were performed using fastp (v0.23.2) (Chen *et al.,* 2018). After fastp processing, only reads with a quality score ≥Q30 were retained; reads shorter than 30 bp or failing quality filters were discarded. The remaining trimmed reads for each sample were carried forward to alignment.

Read Alignment and Gene Counting

Trimmed reads were aligned to the domestic dog reference genome (*Canis lupus familiaris*, CanFam3.1) using HISAT2 (v2.2.1) (Kim *et al.,* 2019). We built a HISAT2 index from the CanFam3.1 reference genome sequence. Each sample's paired-end reads allowed a maximum of 2 mismatches per read and retained up to 10 alignments for reads mapping to repetitive regions.

To quantify gene expression, we employed HTSeq-count (Anders *et al.,* 2015) to count reads mapping to each gene. We used *Canis lupus familiaris* gene annotation (CanFam3.1) to define gene features. Raw counts were then imported into DESeq2 in R for normalization and differential expression analysis.

Differential Expression Analysis

Differential expression between Day 28 (regressed) and Day 0 (initial) tumours was assessed using the DESeq2 Bioconductor package (Love *et al.,* 2014). We created a DESeq2 dataset with a paired design, specifying the dog ID as a blocking factor (since each dog contributes a day 0 and day 28 sample) and condition (Day28 vs Day0) as the main effect. DEGs were identified based on an adjusted p-value threshold of 0.05. We also applied a fold-change threshold, flagging genes with $\log_2$ fold-change ≥ 1. DESeq2 outputs included the list of DEGs with $\log_2$ fold-changes (Day28/Day0) and associated statistics. In addition to the statistical test, we performed exploratory data analysis: variance-stabilizing transformation (VST) was applied to counts for PCA and clustering and sample distances were examined to verify the Day 0 vs Day 28 separation. All analyses were conducted in R (v4.1) using RStudio. Figures (PCA plot, heatmap, volcano plot) were generated in R with ggplot2 and Complex Heatmap packages, and finalized.

Results and Discussion

Sequencing Quality and Alignment Summary

Each sample's total reads (in millions), duplication percentage, GC-content percentages are shown in Table 2. All samples have high sequencing depth and passed quality criteria after fastp filtering.

Table 2: FastQC summary of raw read data and post-filtering results for each sample

Sample ID	Total Sequences (Million)	Duplication (%)	GC (%)
ERR2044811_2	44.34	47.87	43
ERR2044812_1	43.92	49.75	44
ERR2044812_2	43.92	50.93	44
ERR2044813_1	38.14	48.04	44
ERR2044813_2	38.14	50.23	44
ERR2044814_1	40.82	52.74	50
ERR2044814_2	40.82	52.17	51
ERR2044815_1	40.81	47.77	48
ERR2044815_2	40.81	48	48
ERR2044816_1	39.46	47.14	49
ERR2044816_2	39.46	52.31	50
ERR2044817_1	42.71	47.47	48
ERR2044817_2	42.71	47.87	48
ERR2044818_1	40.65	47.8	51
ERR2044818_2	40.65	50.93	51
ERR2044819_1	15.91	48.89	38
ERR2044819_2	15.91	48.85	40
ERR2044820_1	44.75	53	51
ERR2044820_2	44.75	49.57	52
ERR2044821_1	41.75	50.36	48
ERR2044821_2	41.75	48.6	48
ERR2044822_1	45.88	47.17	52
ERR2044822_2	45.88	51.77	53

All 12 RNA-Seq libraries produced approximately 30-80 million raw reads. FastQC quality profiles showed Phred scores well above 30 across all sequencing cycles. The average GC content of the reads was approximately 49%, in line with expectations for canine coding transcripts (which typically range between ~48-52% GC). fastp's trimming results showed minimal adapter presence (<0.1% of reads trimmed for adapter) and only small fractions of low-quality bases removed, suggesting that the data were already clean. A moderate proportion of duplicate reads were observed (FastQC reported that roughly 25-35% of reads per library were duplicates on average). This duplication level is considered normal for deep mRNA-Seq, where highly expressed genes lead to repeated sequencing of the same transcripts. Following deduplication, the loss of reads was estimated at under 30% for all samples, which FastQC flags as a warning but is acceptable for transcriptomic studies. All samples were retained, as each met the minimum quality and read depth thresholds for downstream analysis.

Alignment of reads to the canine reference genome was highly successful. Using HISAT2, we achieved an average overall alignment rate of ~96.4% across all samples. Table 3 summarizes the alignment metrics for representative samples. The majority of reads aligned in proper pairs to unique genomic locations. For example, in one Day 0 sample, 96.7% of reads mapped (91.2% were mapped uniquely, 5.5% to multiple sites), while in its paired Day 28 sample 95.8% mapped (89.0% were uniquely mapped, 6.8% were multi-mapped). Such high unique mapping rates indicate that the CanFam3.1 genome provides comprehensive reference coverage for the expressed transcripts in CTVT. The aligned read data exhibited consistent insert size distributions (200±20 bp) and splice junction mappings that corresponded to known exon-intron boundaries, confirming that the libraries were of good quality. Overall, the QC and alignment results established a solid foundation for reliable differential expression analysis between conditions.

Table 3: Summary of read alignment metrics per sample using HISAT2

Sample ID	Aligned (%)	Uniquely Aligned (%)
ERR2044811	87.54	84.70
ERR2044812	87.11	84.57
ERR2044813	88.96	85.34
ERR2044814	88.91	81.77
ERR2044815	92.99	86.62
ERR2044816	82.85	75.18
ERR2044817	89.28	83.44
ERR2044818	82.37	74.72
ERR2044819	80.64	76.21
ERR2044820	90.23	81.84
ERR2044821	89.46	81.88
ERR2044822	81.27	71.00

Global Gene Expression Patterns Distinguish Regressed vs. Progressive Tumours

To explore overall expression differences, we first performed an unsupervised PCA on the variance-stabilized expression data. The PCA revealed a strong separation of samples by treatment status along the first principal component (PC1).

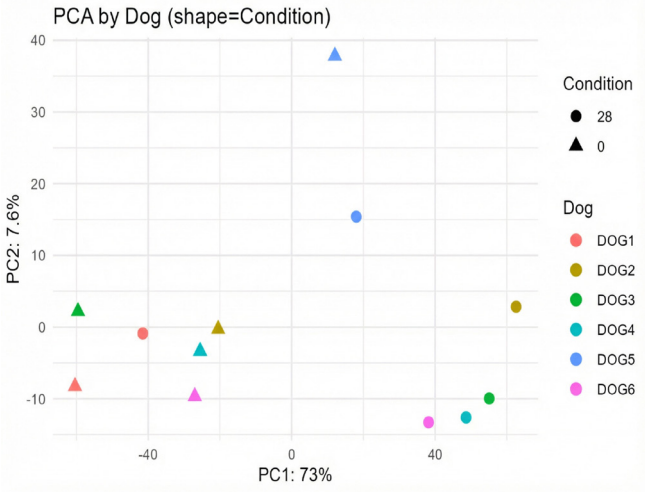

Fig. 1: Principal component analysis (PCA) of variance-stabilized RNA-Seq expression data from paired CTVT samples (Each point represents one sample, colored by individual dog (Dog1-Dog6) and shaped by treatment time point (Day 0 = triangle, Day 28 = circle). Biological replicates from the same subject cluster closer together, while treatment status separates along PC1).

The PCA plot (Fig. 1) illustrates the global gene expression differences between pre-treatment and post-treatment samples. Each point represents a tumour biopsy sample; shapes denote timepoint (▲ = Day 0, ● = Day 28) and colors represent individual dogs (Dog1-Dog6). Principal component 1 (PC1), accounting for 73% of the total variance, separates samples primarily by treatment status, with Day 0 (pre-treatment) samples clustering on the left and Day 28 (post-treatment) samples on the right. PC2 (7.6% variance) reflects within-group variation. For each dog, paired samples (same color) show directional movement from triangle to circle along PC1, suggesting consistent transcriptional changes within individual dogs after vincristine treatment. The separation along PC1 indicates a strong gene expression shift associated with tumour regression, while the tighter clustering of Day 28 samples suggests more homogeneous expression profiles during the regressed state, potentially driven by shared immune response activation. In contrast, Day 0 tumours are more dispersed, reflecting inter-tumour heterogeneity before therapy.

We next examined global expression patterns using a heatmap of the top 500 most variably expressed genes across all samples. Variance-stabilized counts were mean-centered and scaled for each gene to allow relative comparison of expression levels (Fig. 2).

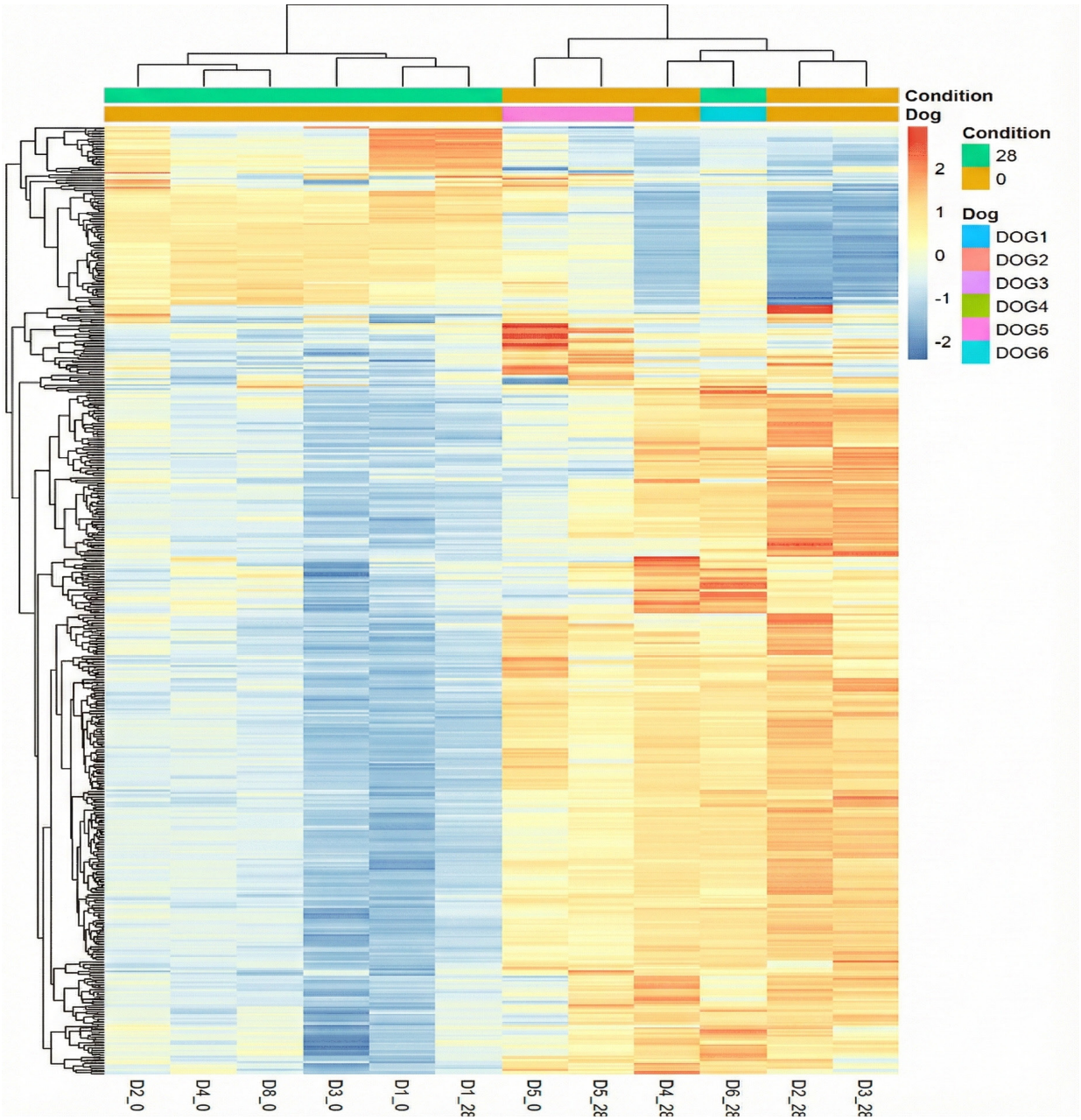

Fig. 2: Heatmap of the Top 500 most variably expressed genes in CTVT samples (Each column represents a sample, grouped by individual dog (Dog1-Dog6) and timepoint (0 = Day 0, 28 = Day 28), and each row represents a gene. Color intensity (blue = low, red = high) reflects relative expression values, which have been mean-centered and scaled per gene.

The heatmap reveals two major transcriptional clusters: one for Day 0 (progressive tumours) and another for Day 28 (regressed tumours), indicating a strong effect of treatment status on gene expression. Within each cluster, samples tend to pair by dog, but the dominant signal reflects the pre- vs. post-treatment state. The heatmap does not display specific gene names, genes contributing to the separation include immune response-related transcripts such as *CXCL10* and *CCL5*, which were upregulated at Day 28, and proliferation markers like *MKI67* and *TOP2A*, which were downregulated after treatment. These patterns support a transition from active tumour growth to immune-mediated regression following vincristine therapy.

Differential Expression Analysis (Day 28 vs Day 0)

Applying DESeq2's statistical analysis, we identified 2,873 DEGs between Day 28 (Regressed tumours) and Day 0 (progressive tumours) using an adjusted p value <0.05. Among these, 1763 genes were upregulated (*i.e.,* higher expression at Day 28) and 1109 were downregulated (lower expression at Day 28), based on a minimum absolute $\log_2$ fold change of 1.5 (equivalent to a fold change ≥ 2). The global expression profile is summarized in a volcano plot (Fig. 3), where significantly upregulated and downregulated genes are visually separated by fold change and statistical significance.

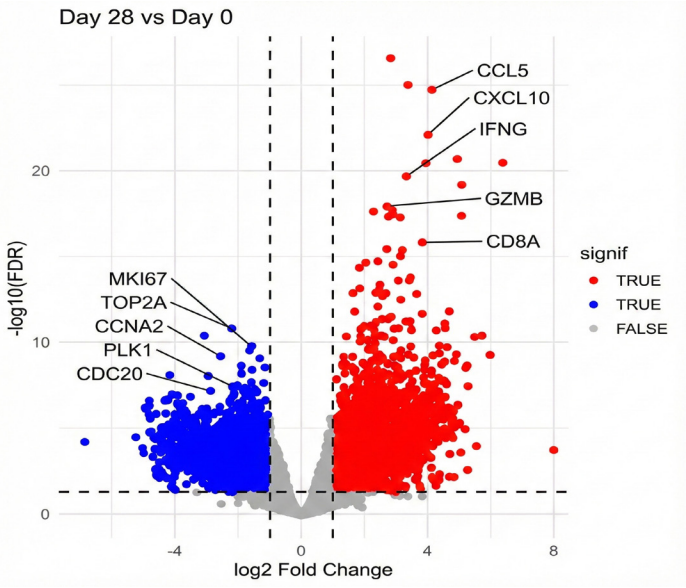

Fig. 3: Volcano plot of differential gene expression results for Day 28 vs Day 0 {Each point represents a gene (out of ~15,325 tested). The x-axis is the $\log_2$ fold change (Day 28 / Day 0), and the y-axis is the $-\log_{10}$ adjusted p-value. Red points indicate significantly upregulated genes (adj p< 0.05), blue points indicate significantly downregulated genes, and gray points are non-significant. Select significant genes are labeled for illustration).

In Volcano plot of differential gene expression between regressed (Day 28) and initial (Day 0) CTVT tumours, each point represents a gene; the *x*-axis is the $\log_2$ fold-change (Day28/Day0) and the *y*-axis is the $-\log_{10}$ adjusted p-value. Genes with significant differential expression (adjusted p-value <0.05) are highlighted in color: red points indicate significantly upregulated genes in Day 28 (positive $\log_2$ fold change) and blue points indicate significantly downregulated genes in Day 28 (negative $\log_2$ foldchange). Non-significant genes are shown in gray. The plot shows a broad pattern of gene expression changes. A substantial number of genes (red cluster on right) are upregulated in regressed tumours, with $\log_2$ fold-changes up to ~+8 (256-fold). Conversely, numerous genes (blue cluster on left) are downregulated in regressed tumours, with $\log_2$ fold-changes down to approximately -5 (about 1/32nd of day 0 levels). The horizontal dashed line marks adjusted p value = 0.05; genes above this line are statistically significant. Notably, *CCL5* ($\log_2$ FC ≈ +4.1, $-\log_{10}$ (padj) ~25) and *CXCL10* ($\log_2$ FC ≈ +3.9, $-\log_{10}$ (padj) ~21.5) are among the significantly upregulated genes in Day 28 tumours. These chemokines show large expression increases in regressed tumours, indicating an active immune chemoattractant environment. On the downregulated side, *MKI67* ($\log_2$ FC ≈ -1.2, $-\log_{10}$ (padj) ~10) and *TOP2A* ($\log_2$ FC ≈ -1.8, $-\log_{10}$ (padj) ~15) stand out as significantly repressed in regressed tumours. Both are proliferation markers, so their suppression reflects reduced tumour cell division. This volcano plot highlights the central biological theme: regressing CTVT tumours display enhanced immune/inflammatory gene expression and suppressed cell cycle activity following vincristine treatment.

Key Differentially Expressed Genes and Biological Interpretation

The transcriptional profile of regressed (Day 28) CTVT tumours reflects strong immune activation and suppression of cell proliferation hallmarks of effective tumour control. Among the most upregulated genes, CCL5 and CXCL10 stand out as key chemokines that recruit T cells, NK cells and monocytes into the tumour microenvironment. Prior work has confirmed their involvement in anti-tumour immunity across various cancers (Tokunaga *et al*., 2018; House *et al.,* 2020) and specifically in CTVT regression (Frampton *et al.,* 2018). The elevated expression of *IFNG, CXCL9,* and *GZMB* further supports a Th1-skewed immune response, as seen in regressing tumours of both human and canine models (Moro *et al.,* 2010; Murchison *et al*., 2014).

Upregulation of antigen presentation genes, including DLA-DRA and DLA-DQA1 (canine MHC II equivalents) reflects enhanced immune visibility of tumor cells. This is consistent with regression-phase histology studies that show heavy lymphocytic infiltration and increased MHC expression (Siddle and Kaufman, 2015; Magi *et al.,* 2017). Interestingly, some studies have also observed paradoxical MHC downregulation in late regression (Amaral *et al.,* 2019), potentially due to tumour clearance or immunoediting.

Conclusion

Taken together, our findings reveal a dual regulatory axis in regressed CTVT: immune activation driving tumour clearance, and growth suppression halting further proliferation. This transcriptomic signature mirrors canonical patterns seen in successful cancer immunotherapy, making CTVT a valuable natural model for studying tumour-immune interactions.

Acknowledgments

The authors thank the original data generators, particularly Frampton *et al.* and collaborators, for making the CTVT RNA-Seq dataset publicly available. We also acknowledge the NCBI Sequence Read Archive for providing access to the sequencing data. We thank colleagues in our department for valuable discussions and academic support.

References

Amaral, M.A., Nunes, F.C., Amorim, R.L., & Dagli, M.L.Z. (2019). Evaluation of MHC class II expression in canine transmissible venereal tumour during regression and progression phases. *Veterinary Comparative Oncology, 17,* 484.

Anders, S., Pyl, P.T., & Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31,* 166.

Belov, K., Murchison, E.P., & Kaufman, J. (2015). The evolution and genetic basis of transmissible cancers in dogs and Tasmanian devils. *Veterinary Pathology, 52,* 786.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34,* 884.

Chiang, S.Y., Chiou, Y.Y., Huang, H.H., Chien, M.S., & Lin, C.T. (2013). Analysis of gene expression profiles in canine transmissible venereal tumour (CTVT) regression. *Veterinary Immunology & Immunopathology, 153,* 52.

Frampton, D., Schwenzer, H., Marino, G., Sarker, S.J., & Murchison, E.P. (2018). Molecular profiling of regression in canine transmissible venereal tumour reveals immune-mediated transcriptional programs and antigenic heterogeneity. *Genome Biology, 19,* 203.

House, I.G., Savas, P., Lai, J., Chen, A.X.Y., Oliver, A.J., Teo, Z.L., Todd, K.L., Henderson, M.A., Giuffrida, L., Petley, E.V., Lim, E., Fear, V.S., Forbes, C.A., & Westwood, J.A. (2020). Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clinical Cancer Research, 26,* 487.

Kim, D., Paggi, J.M., Park, C., Bennett, C., & Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Natural Biotechnology, 37,* 907.

Love, M.I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15,* 550.

Magi, G.E., Amorim, R.L., Dagli, M.L.Z., & Amorim, R.M. (2017). CD3+ T lymphocyte infiltration and apoptosis in spontaneous canine transmissible venereal tumour during regression phase. *Research in Veterinary Science, 112,* 89.

Moro, J.V., Sanches, O.C., Huppes, R.R., Amorim, R.L., Mota, L.S.L., Silveira, J.A.S., Fighera, R.A., Kommers, G.D., & Dagli, M.L.Z. (2010). Immunohistochemical study of tumour cell proliferation and apoptosis in canine transmissible venereal tumour. *Pesquisa Veterinária Brasileira, 30,* 559.

Murchison, E.P., Wedge, D.C., Alexandrov, L.B., Fu, B., Martincorena, I., Ning, Z., Tubio, J.M.C., Werner, E.I., Allen, J., De Nardi, A.B., Donelan, E.M., Marino, G., Fassati, A., Campbell, P.J., Yang, F., Burt, A., Weiss, R.A., & Stratton, M.R. (2014). Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science, 343*(6169), 437-440.

Siddle, H.V., & Kaufman, J. (2015). Immunology of naturally transmissible tumours. *Immunology, 144,* 11.

Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., McSkane, M., Baba, H., & Lenz, H.J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. *Cancer Treatment and Reviews, 63,* 40.