HHS Public Access

Author manuscript

Vet Comp Oncol. Author manuscript; available in PMC 2025 May 10.

Published in final edited form as:

Vet Comp Oncol. 2023 December ; 21(4): 565–577. doi:10.1111/vco.12935.

Leading the pack: Best practices in comparative canine cancer genomics to inform human oncology

Cheryl A. London[1], **Heather Gardner**[1], **Shaying Zhao**[2], **Deborah W. Knapp**[3], **Sagar M. Utturkar**[4], **Dawn L. Duval**[5], **Melissa R. Chambers**[6], **Elaine Ostrander**[7], **Jeffrey M. Trent**[8], **Gina Kuffel**[9]

[1]Cummings School of Veterinary Medicine, Tufts University, North Grafton, Massachusetts, USA

[2]University of Georgia Cancer Center, University of Georgia, Athens, Georgia, USA

[3]College of Veterinary Medicine, Purdue University, West Lafayette, Indiana, USA

[4]Purdue Institute for Cancer Research, Purdue University, West Lafayette, Indiana, USA

[5]College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, Colorado, USA

[6]School of Medicine, University of Alabama, Birmingham, Alabama, USA

[7]Cancer Genetics and Comparative Genomics Branch, National Cancer Institute, Bethesda, Maryland, USA

[8]Translational Genomics Research Institute, Phoenix, Arizona, USA

[9]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

Abstract

Pet dogs develop spontaneous cancers at a rate estimated to be five times higher than that of humans, providing a unique opportunity to study disease biology and evaluate novel therapeutic strategies in a model system that possesses an intact immune system and mirrors key aspects of human cancer biology. Despite decades of interest, effective utilization of pet dog cancers has been hindered by a limited repertoire of necessary cellular and molecular reagents for both in vitro and in vivo studies, as well as a dearth of information regarding the genomic landscape of these cancers. Recently, many of these critical gaps have been addressed through the generation of a highly annotated canine reference genome, the creation of several tools necessary for multi-omic analysis of canine tumours, and the development of a centralized repository for key genomic and associated clinical information from canine cancer patients, the Integrated Canine Data Commons. Together, these advances have catalysed multidisciplinary efforts designed to integrate the study of pet dog cancers more effectively into the translational continuum, with the ultimate goal of

Correspondence: Cheryl A. London, Cummings School of Veterinary Medicine, Tufts University, 200 Westboro Road, North Grafton, MA 01536, USA. cheryl.london@tufts.edu.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

improving human outcomes. The current review summarizes this recent progress and provides a guide to resources and tools available for comparative study of pet dog cancers.

Keywords

canine genome; data analysis; tumour genomics

1 | INTRODUCTION

Mouse models of cancer have been instrumental for dissecting biological processes that underlie tumorigenesis. However, these often lack key features of spontaneous disease including treatment-induced resistance and multi-focal metastases that sometimes occurs a decade or longer post therapy. More recently, genetically engineered mice and humanized models have supported interrogation of novel therapeutic approaches in settings that more closely recapitulate human cancers.[1,2] Despite these improvements, the complexities of cancer, especially with respect to the heterogeneity of response to multi-modal interventions typically used in humans, cannot be fully dissected utilizing these models. This is particularly relevant in the context of combination strategies that incorporate newer approaches such as immune modulation, among others. Consequently, failure rates for oncology therapeutic development remain over 80%, and some types of cancer, such as sarcomas, have seen little improvement in outcomes for over 3–4 decades.[3–5]

Tumours that develop spontaneously in pet dogs due to normal aging and mutations secondary to lifetime exposures share many characteristics with human cancers including histologic appearance, biological behaviour, molecular drivers, heterogeneity, therapeutic response, acquired resistance/recurrence, and patterns of metastasis.[6] Importantly, the annual incidence of cancer in pet dogs is estimated to be up to five times higher than that of humans (1000–2500 vs. 500 new cases/100 000, respectively), with over 4 million new canine cancer diagnoses each year in the United States alone.[6–9] Interest in leveraging pet dogs with spontaneous cancer to facilitate translational research has grown substantially over the past decade. As with humans, pet dogs are immunocompetent and possess a spectrum of co-morbidities that influence disease severity, immune response, tolerance to therapeutic interventions and response and resistance to treatment. Additionally, pet dogs can be followed longitudinally for years, large numbers of samples can be readily collected for biomarker interrogation, and disease can be monitored using standard methodologies (i.e., CT, MRI).[6,10] Studies in pet dogs with cancer have provided key information on toxicity, target validation, dosing and efficacy for several FDA approved human drugs.[11–14] Ongoing work in dogs spans several areas of research such as aging and cancer (NIH funded Dog Aging Project[15]), immuno-oncology (NCI Cancer Moonshot's funded Premedical Cancer Immunotherapy Network Canine Trials [PRECINCT])[16,17] and disease specific initiatives (European Canine Lymphoma Network). Additionally, the National Academies recently held a workshop on the role of companion animals as sentinels for predicting the effects of environmental exposures on aging and cancer susceptibility in humans.[18] Together, these initiatives underscore the breadth of work focused on enhancing ties between veterinary and human cancer research.

With the expanding utilization of pet dogs in preclinical cancer research studies, it has become equally important that tools and resources necessary for cellular, molecular, immunologic and genomic analysis of canine cancers be comparable to those routinely used for human (and mouse) cancer work. Toward that end, several investments have been made by the NIH, non-profit foundations and biomedical industry entities to rapidly grow capacities, all of which were markedly facilitiated by the 2015 National Academies Workshop 'The Role of Clinical Studies for Pets with Naturally Occurring Tumors in Translational Cancer Research'.[19] Together, these inspired an acceleration of efforts to generate the technologies and information critical for effective utilization of pet dogs as a model system in the context of comparative cancer studies, especially those that integrate novel immune-oncology strategies.

2 | CANINE GENOMES

Because dogs were domesticated to become human companions over thousands of years, they developed similar traits through the process of convergent evolution.[20] Indeed, genomic analysis of phylogenetic relationships provides evidence for a primate-carnivore clade with the exclusion of rodents.[21] Beyond these parallels, several dog breeds are predisposed to the development of specific types of cancers, creating an opportunity to identify genetic drivers that contribute to inherited cancer risk in both dogs and man. Such studies require the availability of a high-quality reference genome to assess the potential relevance of germline variants potentially associated with certain cancers, and to determine the functional consequences of both coding and non-coding variants identified in canine tumour cells. Currently, there are seven carefully curated, commonly used and broadly available versions of the canine genome available for public use, detailed in Table 1.

Recently, the canine research community endorsed use of the reference canine genome created from Mischka, a German Shepherd Dog (UU_Cfam_GSD_1.0, also known as canFam4).[22] This assembly was undertaken using long read technology, resulting in comparatively small numbers of gaps, long scaffolds and high levels of completion of conserved single-copy genes. Indeed, chromosome length, scaffolds and contiguity are increased 55-fold over the genome build previously used by the community, canFam3.1. Annotation incorporating both new and existing long and short read RNA-seq, miRNA-seq and ATAC-seq, reveals that existing gaps in canFam3.1 contain previously unknown functional elements, including promoters, genes and non-coding RNAs.

Beyond the validated canine reference genomes, over 2000 additional genomes have been publicly sequenced by individual laboratories and the data are available in the Short Read Archive (SRA) which are increasingly being aligned to canFam4. Despite these improvements, understanding the landscape of normal breed specific variants (i.e., single nucleotide polymorphisms [SNPs] that are associated with normal dog genomes within a breed) remains a challenge. Existing data would suggest that identification of breeds through SNPs is unlikely to be accomplished using one or a small number of SNPs, but rather, a pattern of SNPs spanning the genome will define each breed.[23] To address this and facilitate mapping of genome encoded traits, a based variant call file (VCF) was constructed to create a framework for classification of large numbers of SNPs,[24] which has undergone multiple

iterations to expand the breadth of known germline variants. Because the National Center for Biotechnology Information Single Nucleotide Polymorphism database (dbSNP) is a mosaic database with both germline and somatic variants, and has not supported dog data since 2018, all new canine sequence data has been loaded to the SRA including VCF variant files, the latter of which are particularly valuable to researchers. The reference sequences are loaded in the UCSC browser, as is the annotation.

In addition to individual lab efforts, an ongoing organized project to comprehensively define the landscape of canine genomics, inclusive of all breeds, has being undertaken by the Dog10K Consortium.[25] This group has sequenced 1987 additional genomes from over 200 established dog breeds, with multiple dogs of each breed undergoing whole genome sequencing (WGS), together with mixed breed dogs, village dogs and wild canids. Data are aligned to the German Shepherd Dog reference sequence (canFam4) and called together. In addition to SNPs, analysis of these data includes that of structural variation, mitochondria DNA and diversity across genomes. SNV and SV variants have been deposited to the European Variation Archive and mitochondrial genomes are submitted to GenBank. Variant files and associated annotations are available at https://kiddlabshare.med.umich.edu/dog10K/ and raw sequence data are available under accessions PRJNA648123 and PRJNA188158.

3 | CANINE CANCER DATA REPOSITORIES

There are now several data repositories that will house information relevant to canine cancer studies, including genomic information, imaging data and clinical annotations. These are described below.

3.1 | Integrated Canine Data Commons (ICDC)

NCI's Division of Cancer Treatment and Diagnosis (DCTD) appointed the Frederick National Laboratory for Cancer Research (FNLCR) to build the Integrated Canine Data Commons (ICDC) (https://caninecommons.cancer.gov/#/), a publicly available cloud-based repository of spontaneously arising canine cancer data. The ICDC was established to further research on human cancers by enabling comparative analysis with canine cancer and to make canine cancer data more findable, accessible, interoperable and reusable (FAIR). Data in the ICDC are sourced from multiple programs and studies, focused primarily on data generated from pet dogs with cancer. The data are harmonized into an integrated data model and then made available to the research community. The ICDC accepts a wide range of data and metadata types including, descriptive case level information, genetic data (including WGS, whole exome sequencing [WES], RNA sequencing [RNA-seq], among others), clinical data (treatment information, patient outcomes, adverse events), and imaging (MRI, CT, etc.). Currently, the ICDC is home to 35 TB of data, including 3 different scientific programs, comprising 11 studies, 678 cases and 937 samples. Importantly, ICDC provides html links for studies that have other data types available for canine cases that are hosted on other systems, as well as links to the human data nodes in the NCI Cancer Research Data Commons (CRDC).

3.2 | The Cancer Imaging Archive (TCIA)

The Cancer Imaging Archive is part of the NCI's Cancer Imaging Program in the Division of Cancer Treatment and Diagnosis (DCTD). TCIA is a service that effectively redacts and hosts a large archive of medical images of cancer that can then be made publicly accessible through a data portal or through an application programming interface. Currently, 81 canine subjects have images hosted in TCIA (https://cancerimagingarchive.net/), including baseline pathologic, MRI, and genomic features of treatment-naive naturally occurring canine glioma available for direct download. Radiology imaging data are available for 57 of these subjects, and pathology imaging data for 78 subjects. These were collected from 5 different veterinary institutions' patient populations. Clinical MRIs are available prior to treatment, and surgical biopsy and/or tissues collected at necropsy are provided as H&E-stained sections. The tumours have been classified with a harmonized scheme developed by veterinary and physician neuropathologists. The corresponding multi-omic and structured clinical metadata are accessible through the ICDC. The first dataset of its kind to comprehensively describe and report the clinical, pathologic, imaging, and genomic landscape of naturally occurring canine glioma, it provides the necessary context to define this canine tumour as a model for its human counterpart(s).

3.3 | Imaging Data Commons (IDC)

The Imaging Data Commons is a node within the broader NCI Cancer Research Data Commons (CRDC) infrastructure. IDC (https://imaging.datacommons.cancer.gov/) is supported by the Cancer Moonshot with the goal of enabling accessibility to de-identified imaging data to support integrated analyses with other non-imaging data types utilizing the CRDC Cloud Resources. The IDC boasts a powerful user interface for exploring available data collections and supports the integration of a DICOM viewer to inspect images on demand. The images available from the 57 subjects hosted in the TCIA can be viewed as a collection directly from within the IDC upon selection.

4 | CLOUD RESOURCES FOR CANINE DATA ANALYSIS

4.1 | Seven Bridges Cancer Genomics Cloud (SB CGC)

A main goal of the CRDC is to co-locate data with cloud computing resources to accelerate data-driven discovery across diverse sets of information. The Cancer Genomics Cloud (CGC) (https://cgc.sbgenomics.com/) is powered by Seven Bridges and serves as a secure cloud-based analytic platform for cancer research funded by the NCI. New users (including those employed by the federal government) are provided with $300 of compute and storage credits upon registration. Users can access many large NCI datasets across multiple fields directly or data can easily be exported from any of the nodes with the CRDC as a CGC compatible manifest. The CGC hosts a robust library of out of the box peer reviewed bioinformatic pipelines and tools. Additionally, users can bring in their own pipelines or leverage or modify existing tools using their development kit that wrap tools using the Common Workflow Language (CWL). Users can easily build a cohort of interest employing the 'Explore' page in the ICDC and download a manifest of case IDs that can be readily imported into the CGC. This interoperability between systems precludes the need to download ICDC files locally; instead, the manifest provides all instructions necessary for the

CGC to access files directly from ICDC cloud storage, along with the file metadata and the cases/samples from which they were derived. Once files are accessible from within the CGC they can be used as inputs for analysis workflows.

4.2 | Galaxy

Galaxy (https://usegalaxy.org/) is an open source, web-based platform for data intensive biomedical research. Galaxy allows users to bring their own genomic data to serve as inputs for bioinformatic tools and workflows. Galaxy provides a digestible point and click interface, allowing researchers to focus on analysis results instead of coding semantics, thereby lowering the activation energy required to get started with bioinformatics. The free public web server version of Galaxy is supported by the Galaxy project and includes many bioinformatic tools that are broadly useful. Users can create an account, save analysis histories, workflows, and even datasets on the public server. Further, pipelines and analysis workflows can easily be shared across collaborating partners.

4.3 | Terra

Terra (https://terra.bio/) is a cloud-based scalable platform for biomedical research developed by the Broad Institute and powered by the Google Cloud Platform (GCP). The platform is modular, open source, community-driven, and standards-based with the goal of facilitating collaboration among researchers and tool developers across the globe to accelerate scientific discovery. Terra leverages individual workspaces where users can bring their own data and employ popular cloud-based analysis tools. The users can securely share data, code, and analysis results with others in Terra, which has helped in the adaption of analytic pipelines for canine genomics. Registration on Terra is free using any Google managed identity and Google provides $300 in cloud credits for new users to try the platform before committing their own funds.

5 | ANALYTIC TOOLS FOR CANINE GENOMICS

5.1 | Standard pipelines

Standard pipelines are used for routine analyses such as mutation discovery with genomic sequencing reads (e.g., WGS or WES) and gene expression quantification with RNA-seq reads.[26–28] Typical pipelines include sequence quality control, alignment to the reference genome, mutation calling and annotation, and transcript/gene assembly and expression level quantification, with some of the commonly used software tools listed in Table 2. For sequence quality control, sequence reads are typically cleaned up with Trimmomatic, which removes adapter sequences and low-quality bases, and examined by FastQC. The cleaned sequences are aligned to a canine reference genome such as canFam3 or canFam4 (Table 1) using BWA[29,30] for genomic reads (e.g., WGS or WES) and HISAT2 or Tophat[31] for RNA-seq reads. The alignment data are then used to further assess sequence quality by determining mapping rate and quality as described.[32] The read alignments of samples that pass the quality control are further processed with Samtools[30] for sorting or other manipulations, Picard for marking and removing duplicates, and GATK[33] for insertion/deletion (indel) realignment and base quality recalibration. Then, the processed alignment data are ready for mutation calling (e.g., WES or WGS) or expression quantification (RNA-

seq). Mutations are often called with GATK HaplotypCaller (germline), MuTect2 (somatic), Strelka (germline and somatic), and Manta (germline and somatic). Mutations are filtered for sequencing errors and other artifacts,[32] and annotated with tools such as ANNOVAR and Funcotator. Gene/transcript assembly and expression quantification are performed with tools such as Cufflinks, with or without a gene annotation reference file.

When performing these analyses, parameter selection for the tools is critical and often requires optimization, for example, using already known germline or somatic mutations to improve mutation calling. Except when using BWA for mapping, each step (e.g., mutation calling) should be conducted using two or more software tools to assess the concordance of calls. The choice of specific reference genome and its gene annotations also matters. For example, indel realignment can present challenges when using canFam3.1 as the reference genome (created in 2005) is incomplete with many gaps. As noted above, there are several new reference genomes, with the canFam4 providing the most complete coverage, particularly of non-coding regions.[22] However, it is important to note that currently, gene annotations associated with canFam3.1 (which can be downloaded from ensembl.org) are more comprehensive than those linked to canFam4. For somatic mutation calling, filtering germline mutations among dogs is required and hence choosing a reference genome with the most accurate and highly curated set of SNPs known is critical. Resources such as Fido-SNP[34] will also help mutation annotation. As such, several of the cloud resources noted in the previous section enable community problem solving when adapting human analytic pipelines to dog analytics.

With respect to WES, the first canine specific design based on the canFam3.1 genome build and capturing 53 MB was published in 2014.[35] Additional custom canFam3 designs have since been generated and synthesized by Roche (Indianapolis, IN) for WES studies. In addition, Agilent (Santa Clara, CA) has made two canine whole exome captures available to the canine research community, Version 1 based on CanFam2 (part number 5190-5452, 53.59 MB) and Version 2 based on canFam3.1 (part number 5191-6700, 43.45 MB). More recently, Twist Bioscience (South San Francisco, CA) in collaboration with the Karlsson Lab at the Broad Institute has developed a commercially available 40.5 MB canine exome capture based on canFam3.1.

Finally, while several CNV callers as well as specified toolkits that include CNV analysis are listed in Table 2, additional approaches have been employed to broaden the scope of data interrogation, including those used to evaluate clonality and tumour heterogeneity. It is important to note, however, that descriptions of these approaches are generally limited to individual labs. For example, GISTIC2 does not support use in the canine genome without source code modification: https://github.com/sbamin/canine_gistic2. Lastly, methods have been developed to assess the intra- and inter-tumoural heterogeneity of canine cancers (such as osteosarcoma and hemangiosarcoma), although such pipelines have generally been individually curated to address the specific data sets.[36,37]

5.2 | MHCI genotyping

A human individual expresses 3–6 alleles encoded by three classical MHC Class I genes HLA-A, HLA-B, and HLA-C. The syntenic canine MHC Class I region encodes genes

including DLA-88, DLA-12, and DLA-64. The MHC Class I genes are highly polymeric, especially in exons 2 and 3, which encode the antigen-binding pocket. As such, these are among the most hypervariable regions in the genome. Accurate determination of the ~550 bp sequence of exons 2 and 3, typically referred to as MHC genotyping, is used to identify the Class I alleles in a single individual. Extensive work has been undertaken to define the landscape of human MHC Class I (HLA-I) alleles. The IPD-IMGT/HLA Database (https://www.ebi.ac.uk/ipd/imgt/hla/) provides a specialist database for sequences of the human MHC and includes official sequences curated by the World Health Organization Nomenclature Committee For Factors of the HLA System. As of March 2022, the IMGT/HLA database[51] contains 23 694 HLA-I alleles. Next generation sequencing (NGS)-based genotyping tools,[52–61] which focus on human genotyping, take advantage of this vast number of known HLA-I alleles. Because so many HLA-I alleles existing in the human population have been precisely characterized, the current HLA-I genotyping tools focus on accurately mapping NGS reads to the known allele reference, and then genotyping the individual by identifying known alleles with the most unambiguously placed NGS reads.

In dogs, the landscape of MHC Class I alleles has not been well characterized. A total of 185 alleles have been classified including DLA-88 (150 alleles), DLA-88L (8 alleles), DAL-12 (20 alleles), DLA-64 (7 alleles) and DLA-79 (8 alleles)[62–66] most of which are included in the IPD-MHC database (https://www.ebi.ac.uk/ipd/mhc/). Given the paucity of information regarding canine DLA-I alleles, the workflow employed by HLA-I genotyping tools[52–61] that relies upon a well curated database of known alleles, does not work for the dog. To address this deficiency, a new software tool has been developed for use in the dog that outperforms currently used approaches for typing new alleles.[67] Named the Kmer-based paired-end read (KPR) de novo assembler and genotyper, this tool assembles paired-end RNA-seq reads (fastq files) from MHC-I regions into contigs, and then genotypes each assembled contig and estimates its expression level.[67] It is written in Python (version 2.7), runs on the Unix/Linux platform and is freely available at Github at https://github.com/ZhaoS-Lab/KPR.git.

5.3 | Breed validation pipeline

As previously noted, the diverse array of more than 300 pure breeds of dog provides a valuable resource for studies of phenotypic and genotypic evolution, identification of genes associated with disease, and interrogation of how the genomic landscape may influence response to stressors (environment, aging, etc.). However, information regarding breed assignment found in published papers, medical records and other databases is typically provided by pet owners, and therefore not necessarily accurate. This is particularly relevant for common breeds like the Golden Retriever, which can appear to be purebred but may actually be mixed with other breeds. Additionally, the rapid growth of ‘doodles’ which are purported to be various purebred dogs mixed with the standard poodle (i.e., Goldendoodle, Labradoodle, etc.) has resulted in several newer mixed breeds that may not really be derived from the cross between two purebred dogs. These issues can be further compounded by human error (i.e., sample mistracking and or breed mis-assignment). Finally, information on breed may simply be missing from medical records/data bases. Together, these problems can

hamper the ability to effectively utilize canine breed-specific disease to interrogate genomic associations.

To address this issue, a breed validation and prediction software tool has been developed that uses WGS or WES data from dogs to identify breed-specific variants.[32] This methodology requires information from greater than 10 dogs per breed for multiple breeds to build a knowledge base regarding variant calls. The software then validates or predicts the breed of a dog using the variant allele frequency (VAF) values of these breed-specific variants. The software has been successfully used to identify 5463 breed-specific germline sequence variants for 10 pure breeds (e.g., Golden Retriever, Rottweiler, Maltese, etc.) and then further credentialled using breed data from an additional 440 dogs.[32] The pipeline, example input files and the manual for canine breed validation and prediction have been deposited in GitHub, free to the public at https://github.com/ZhaoS-Lab/breed_prediction and https://doi.org/10.5281/zenodo.4948044. The ability to accurately predict breed will further enhance genomic studies in dogs where such information is not readily available.

6 | EMERGING TECHNOLOGIES AND RESOURCES

Over the past decade, the ability to critically assess changes in gene expression in cancer cells and their resultant impact on the microenvironment has markedly improved. These technologies have facilitated assessment of tumour cell dynamics at the single cell level, within a tissue, and more recently, through evaluation of circulating components. Importantly, reagents and tools have now been developed that enable use of these approaches in the context of the canine genome. Together, these provide a set of tools to precisely interpret and apply findings from translational studies in dogs with cancer.

6.1 | Single cell omics

Recent technological advances have enabled researchers to define gene and protein expression, DNA methylation, and chromatin accessibility with single cell resolution.[68] Much of this progress has been in the realm of single-cell RNA sequencing (scRNA-seq) which allows simultaneous measurement of gene expression across thousands to millions of cells in each cell population. This has been achieved through several scientific innovations including the: (1) ability to amplify RNA from single cells in quantities sufficient for sequencing, (2) addition of cell-specific barcodes allowing pooling of multiple cells into a single library, (3) use of in-vitro transcription,[69] a linear amplification method that reduces bias toward highly expressed genes and improves sensitivity, and (4) cost-effective solutions achieved by application of microfluidics and lab-automation. There are now several well-established methodologic approaches and commercial platforms available to support scRNA-seq.[70–73] The workflow for a typical scRNA-seq study consists of: sample preparation, individual cell capturing, RNA isolation, reverse transcription of RNA to cDNA, amplification, library preparation, sequencing, and data analyses. Cells from tissues can be collected by laser capture micro-dissection or by the creation of a single-cell suspension from the tumour tissue, which requires mechanical and enzymatic methods that must be optimized in advance. Methods to isolate single cells from the suspension include

limiting dilution, micromanipulation, flow-activated cell sorting (FACS) and more recently, microfluidic methods such as Drop-Seq.[74]

Analysis of sequencing data first requires pre-processing (data demultiplexing, alignment, feature annotation and quantification) to derive a gene-expression matrix ($M(cells) \times M(genes)$). The next steps involve: (1) quality control including determining low-quality (damaged/doublet) cells by count-depth, elevated mitochondrial gene content or dedicated methods like DoubletFinder or equivalent,[75] (2) normalization to remove amplification/count-depth biases, (3) optional data correction to handle batch effects, biological effects, or dropouts via data imputation methods[76] and (4) dimensionality reduction summarizing the data into a few optimal dimensions while preserving the variability and global structure. A variety of tools in support of these tasks are available[77] and summarized in Table 3, along with suggestions for application of these tools to canine data. End-to-end workflows such as Seurat[78] and Scanpy[79] have also been published. Subsequent downstream analytic steps are typically specific for research questions pertinent to the project and often involve differential expression across biological conditions, cell-type identification, understanding cellular trajectories, novel cellular markers, and unique cellular features. A detailed description of data analysis steps[80,81] and a catalogue of available software tools[77] have also been reported elsewhere. It is important to note that scRNA-seq data analysis does not follow a 'one-size fits all' approach, and it is recommended that a cocktail of methods be used to address study-specific biological questions.

Within the realm of canine studies, a number of publications now exist demonstrating that scRNA-seq is feasible and of value to precisely define the transcriptomic landscape of several canine diseases, such as mammary cancer and pulmonary fibrosis.[91–97] Recently, scRNAseq was combined with immune repertoire sequencing to generate a detailed atlas of peripheral blood TCR$\alpha\beta^+$ T cells in healthy dogs.[98] Based on differential gene expression, distinct differentiation states (naïve, effector, memory, exhausted) and lineages (CD4 T helper and regulatory cells, etc.) were identified, many of which had not previously been described in dogs. Several studies are ongoing to define transcriptional signatures associated with tumour cells, components of the tumour microenvironment, and the immune landscape in canine cancers, with the goal of comparing these to comparable signatures in human cancers (Figure 1). This body of work will facilitate more accurate application of data generated from clinical studies performed in dogs with cancer to future human translation.

6.2 | Spatial transcriptomics

Spatial transcriptomics involves whole or targeted partial transcriptomic data, isolated from histologic regions.[99,100] While multiple approaches have been described, the most commonly used commercial assays are made by 10× Genomics (Visium) and NanoString (GeoMx). Both platforms can be processed with formalin-fixed paraffin embedded (FFPE) samples, as well as Optimal Cutting Temperature compound (OCT) embedded frozen samples. The Visium platform leverages a 3′ poly-A tail library and analyses single cells individually, utilizing technology analogous to that employed in single-cell 3′ assay sequencing. While this approach can be successfully implemented in OCT frozen tissue, a species-specific targeted panel is required if FFPE tissues are assayed. The GeoMx platform

requires a probe panel for both frozen and FFPE tissues. For the GeoMx workflow, fresh slides are cut and up to four fluorescent-tagged antibodies are used to identify morphologic regions of interest. A barcoded gene panel is then bound to the slide. After scanning slide images, regions of interest up to 600 microns in size are collected via UV illumination of the selected areas on the slide. Each spot contains multiple cells from a specific region of interest, and gene expression is characterized using short-read sequencing or the nCounter system (see below). While human and murine GeoMx panels can be used in other species, validation and optimization should be undertaken to ensure the panel probes recognize canine transcripts. To facilitate the application of GeoMx to canine cancer, NanoString has created a 1900+ gene canine specific panel (GeoMx Canine Cancer Atlas) that closely mirrors the available human Cancer Transcriptome Atlas, with a slightly broader inclusion of relevant immuno-oncology genes. An additional consideration for spatial transcriptomics is the use of tissue microarrays to support analysis of multiple tissue samples per assay. This process helps minimize the cost of the experiment but requires pre-selection of tissue regions. Furthermore, while 1 mm spot sizes can be successfully used, little room exists to spatially select regions based on fluorescent morphologic antibody binding, as the 1 mm punch size is only slightly larger than the largest 600-micron area of illumination for region collection. The ability to leverage banked FFPE tissue to understand gene expression in the context of the tumour microenvironment sets the stage for further credentialling of canine cancer models to enhance their use in preclinical research.

6.3 | nCounter assay

The nCounter Analysis System was developed by NanoString in the early 2000s, leveraging technology developed at the Institute for Systems Biology.[101] This platform utilizes a relatively straightforward hybridization that permits digital counting of individual RNA molecules without the need for any enzymatic steps. The nCounter system has capacity to detect 800+ unique molecules and is customizable, permitting substantial flexibility to utilize pre-formulated panels or to design new panels for discovery. The workflow entails the use of capture probes and then unique optical barcodes that hybridize to specific targets for detection. The target/probe complexes are then individually counted to give a quantitative assessment of each RNA molecule in the sample analysed. The nCoutner system has flexibility for input of RNA from tissues, fluids and FFPE samples.

Application of nCounter Analysis for comparative and translational work involving pet dogs with spontaneous disease has been facilitated through the development of a canine-specific immuno-oncology panel (Canine IO Panel) with the capacity to profile 800 genes across 47 annotated pathways that overlap with the NanoString Human PanCancer 360 and Immune Panels. The Canine IO Panel was designed to have broad coverage, while also possessing content relevant for the top canine cancers (i.e., lymphoma, osteosarcoma, etc.). There are now several examples demonstrating utility of the Canine IO panel for discovery and/or to credential a spontaneous canine model. For example, the platform was used to interrogate shared inflammatory and skin-specific gene signatures reveal common drivers of discoid lupus erythematosus (DLE) in canines, humans and mice.[102] Interestingly, canine samples recapitulated downregulated skin homeostatic genes observed in human DLE supporting the notion that the canine disease is a relevant translational model. More recently, high intensity

focused ultrasound (HIFU) was used to treat pet dogs with solid tumours and nCounter analysis of biopsy samples demonstrated differential expression of 28 genes associated with T-cell activation following treatment.[103] Several other groups have published gene expression data generated from canine samples using the nCounter platform, and a multitude of studies are ongoing using this tool as part of the PRECINCT consortium.[104–106] It is worth noting that custom panels can be designed to facilitate targeted interrogation. This approach was recently employed to support validation of an antibody panel designed to distinguish lymphoid versus myeloid lineage leukaemias, demonstrating 100% concordance between flow cytometry and gene expression for the lymphoid leukaemias and 80% concordance for myeloid leukaemias.[107]

6.4 | Dependency mapping

The Cancer Dependency Map (DepMap) (https://depmap.org/portal/depmap/) is an ongoing open science project at the Broad Institute working to identify essential genes across hundreds of cancer cell lines. The project combines three basic approaches. The first involves molecular characterization of existing and new cell lines and was known as the Cancer Cell Line Encyclopedia (CCLE). Originally an independent project initiated in 2008 as a collaboration between the Broad and the Genomics Institute of the Novartis Research Foundation, it was merged into the DepMap project in 2018. The CCLE undertook detailed genetic and pharmacologic characterization of a large panel of human cancer models to develop integrated computational analyses that linked pharmacologic vulnerabilities to genomic data, with the goal of translating this information into cancer patient stratification. The second, formerly termed Project Achilles, involves using genome wide loss of function screens (CRISPR, shRNA/RNAi) to create a catalogue of genetic vulnerabilities. Use of this resource has been facilitated by advanced computational modelling to more accurately interrogate the essentiality of individual genes identified through the screens. The last component involves using a unique assay termed PRISM (www.theprismlab.org) to evaluate the activity of anti-cancer agents using a multiplexed cell line platform. The PRISM-generated sensitivity profiles are integrated with genomic information to create predictive models that correlate specific features with drug sensitivities. Data from all three of these efforts is integrated to establish predictive modelling that includes computational models of cancer vulnerabilities and a cancer dependency map. This map incorporates a combination of genetic targets, therapeutic leads and patient stratification.

While there is not yet a DepMap for canine cell lines, efforts are currently underway to develop such a project at the Broad Institute.[108] This will tentatively include WGS across 100 canine cancer cell lines, with 50 lines also undergoing RNA-Seq (personal communication). Recently, canine specific genome wide CRISPR screens have been developed that are at various stages of curation.[109,110] As these were constructed based on canFam3.1, it is likely they will undergo improvements in scope and accuracy of targeting following release of canFam4. In the interim, some canine cancer cell line characterization has already been completed. For example, WES was performed on 33 canine cancer cell lines from 10 cancer types to identify somatic variants that contribute to pathogenesis and therapeutic sensitivity.[111] The authors discovered 501 non-COSMIC genes with PCR variants that functionally annotated with COSMIC genes. They also frequently mapped to

the KEGG MAPK and PI3K-AKT pathways. Twelve of the 33 cell lines were trametinib-sensitive, all 12 exhibited constitutive or serum-activated ERK1/2 phosphorylation, and 8 carried MAPK pathway cancer driver variants: NF1 (2), BRAF (3), N/KRAS (3). Additionally, eight canine osteosarcoma cell lines underwent WGS, confirming a genomic landscape that mirrors that previously defined for canine osteosarcoma using primary tumour biopsies including *TP53* mutations, *MYC* copy number gains, loss of *CDKN2A*, *PTEN*, *DLG2*, *MAGI2*, and *RB1* and structural variants involving *SETD2*, *DLG2*, and *DMD*. Together, these data provide preliminary evidence that a canine DepMap effort will likely yield important data regarding therapeutic vulnerabilities.

6.5 | Liquid biopsy

Liquid biopsy has emerged as a minimally invasive tool for the detection and monitoring of cancer.[112,113] It relies upon the fact that dead/dying cells release DNA and RNA, among other components, that can be detected in body fluids, such as the blood/plasma and/or urine. To date, most efforts have focused on identifying circulating tumour (ct-) DNA from cell-free (cf-) DNA, followed by next-generation sequencing or digital PCR (dPCR). Ultra-low-pass whole genome sequencing (ULPWGS) is commonly performed at a depth of 0.1× and is therefore relatively inexpensive. iChorCNA can then be employed post ULPWGS to generate information on tumour fraction (fraction of ctDNA in total cfDNA) and high-level copy number aberrations. In cfDNA samples with a tumour fraction of at least 10%, WES can be undertaken to achieve increased depth for variant calling using standard tools.[114] However, the limit of detection for ULPWGS is a tumour fraction of 3%–5% and as such, it may be challenging to use this approach for early detection of cancer when the tumour fraction is quite low.[115,116] Sensitivity of detection can be improved using allele specific or drop digital PCR; however, this methodology requires pre-existing knowledge of recurrent single nucleotide variants (SNVs) and short indels. This is the approach utilized with the Cadet BRAF mutation test for canine urothelial carcinomas, leveraging the fact that greater than 80% of these tumours possess a specific BRAF SNV (V595E).[117,118] Additionally, this BRAF mutation is detectable in plasma from dogs with urothelial carcinoma using allele-specific qPCR and the level of *BRAF*-mutated ctDNA in plasma correlates with disease progression.[119] More recently, a commercial canine blood biopsy test also was developed to aid in the diagnosis of cancer in dogs. However, this test has a relatively low sensitivity, particularly for detection of early cancers (i.e., sensitivity of only 20% for detecting cancers less than 5 cm in size).[120] These data emphasize the challenges associated with using liquid biopsy as a screening tool for early detection.

In human oncology, there are now FDA-approved liquid biopsy diagnostics that are typically used as an adjunct to standard diagnostic methodologies (i.e., diagnostic in aid), not as a screening tool. These include the Guardant360 CDx and FoundationOne Liquid CDx tests, among others. The assays rely on an exome capture panel approach to detect specific variants/mutations known to be associated with a range of human cancers such as lung cancer, melanoma, and breast cancer. However, challenges remain with respect to sensitivity for early detection of cancer and for minimal residual disease monitoring. Additionally, the exome capture panel methodology does not work well for tumours with more chaotic tumour genomes and few recurrent mutations, in which the implicit inter-

tumoural heterogeneity may result in an increased incidence of false negative calls. Toward that end, efforts are underway to expand the capacities of liquid biopsy by employing a multi-omics approach integrating information from DNA fragment size, epigenetic analysis of ctDNA, and ctRNA sequencing, among others. Additional considerations to improve liquid biopsy in both dogs and people include the identification of clonal haematopoiesis of indeterminate potential (CHIP), and improved understanding of the mechanisms associated with cfDNA release in the blood.

7 | SUMMARY

Incorporation of studies involving pet dogs with cancer into the preclinical research paradigm has increased substantially over the past decade, driven in part by the goal of leveraging immunologically intact model systems that more closely recapitulate human cancer biology. A major limitation to the use of pet dogs in this context has been the lack of tools, such as canine-specific monoclonal antibodies, and resources necessary for detailed multi-omic interrogation to both validate the model and assess mechanistic drivers of therapeutic response and resistance. The rapid pace of methodologic innovation, along with investments from both the NCI and biomedical entities, have facilitated creation of several new canine specific tools that are truly impacting the depth and breadth of translational canine cancer research. It is likely this effort will continue to grow in parallel with the evolving technologies over the coming decade.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Nadia Lanman, Purdue University, for contributing to and review of the work on single cell omics.

REFERENCES

1. Connolly KA, Fitzgerald B, Damo M, Joshi NS. Novel mouse models for cancer immunology. Ann Rev Cancer Biol. 2022;6(1):269–291. [PubMed: 36875867]
2. Zloza A, Karolina Palucka A, Coussens LM, et al. Workshop on challenges, insights, and future directions for mouse and humanized models in cancer immunology and immunotherapy: a report from the associated programs of the 2016 Annual Meeting for the Society for Immunotherapy of Cancer. J Immunother Cancer. 2017;5(1):77. [PubMed: 28923102]
3. Austin CP. Opportunities and challenges in translational science. Clin Transl Sci. 2021;14(5):1629–1647. [PubMed: 33982407]
4. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. Transl Med Commun. 2019;4(1):18.
5. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. Nat Rev Drug Discov. 2019;18(7):495–496. [PubMed: 31267067]
6. LeBlanc AK, Mazcko CN. Improving human cancer therapy through the evaluation of pet dogs. Nat Rev Cancer. 2020;20(12):727–742. [PubMed: 32934365]
7. Crescio MI, Ru G, Aresu L, et al. The Italian Network of Laboratories for Veterinary Oncology (NILOV) 2.0: improving knowledge on canine tumours. Vet Sci. 2022;9(8);394–411. [PubMed: 36006310]
8. Baioni E, Scanziani E, Vincenti MC, et al. Estimating canine cancer incidence: findings from a population-based tumour registry in northwestern Italy. BMC Vet Res. 2017;13(1):203. [PubMed: 28659149]

9. Sarver AL, Makielski KM, DePauw TA, Schulte AJ, Modiano JF. Increased risk of cancer in dogs and humans: a consequence of recent extension of lifespan beyond evolutionarily determined limitations? Aging and Cancer. 2022;3(1):3–19. [PubMed: 35993010]
10. Momozawa Y. The potential of translational research in dogs in human medicine. Transl Regul Sci. 2019;1(1):31–36.
11. London CA, Hannah AL, Zadovoskaya R, et al. Phase I dose-escalating study of SU11654, a small molecule receptor tyrosine kinase inhibitor, in dogs with spontaneous malignancies. Clin Cancer Res. 2003;9(7):2755–2768. [PubMed: 12855656]
12. Pryer NK, Lee LB, Zadovaskaya R, et al. Proof of target for SU11654: inhibition of KIT phosphorylation in canine mast cell tumors. Clin Cancer Res. 2003;9(15):5729–5734. [PubMed: 14654558]
13. London CA, Bernabe LF, Barnard S, et al. Preclinical evaluation of the novel, orally bioavailable selective inhibitor of nuclear export (SINE) KPT-335 in spontaneous canine cancer: results of a phase I study. PloS One. 2014;9(2):e87585. [PubMed: 24503695]
14. Harrington BK, Gardner HL, Izumi R, et al. Preclinical evaluation of the novel BTK inhibitor acalabrutinib in canine models of B-cell non-Hodgkin lymphoma. PloS One. 2016;11(7):e0159607. [PubMed: 27434128]
15. Creevy KE, Akey JM, Kaeberlein M, et al. An open science study of ageing in companion dogs. Nature. 2022;602(7895):51–57. [PubMed: 35110758]
16. Comparative Oncology Program, National Cancer Institute.
17. Premedical Cancer Immunotherapy Network Canine Trials. precinct.org.
18. The role of companion animals as sentinels for predicting environmental exposure effects on aging and cancer susceptibility in humans: a workshop. Paper Presented at: 2021; Washington, DC.
19. The role of clinical studies for pets with naturally occurring tumors in translational cancer research: Workshop summary; Washington, DC (2015).
20. Wang GD, Zhai W, Yang HC, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. Nat Commun. 2013;4:1860. [PubMed: 23673645]
21. Cannarozzi G, Schneider A, Gonnet G. A phylogenomic study of human, dog, and mouse. PLoS Comput Biol. 2007;3(1):e2. [PubMed: 17206860]
22. Wang C, Wallerman O, Arendt ML, et al. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol. 2021;4(1):185. [PubMed: 33568770]
23. Morrill K, Hekman J, Li X, et al. Ancestry-inclusive dog genomics challenges popular breed stereotypes. Science. 2022;376(6592):eabk0639. [PubMed: 35482869]
24. Plassais J, Kim J, Davis BW, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun. 2019;10(1):1489. [PubMed: 30940804]
25. Ostrander EA, Wang G-D, Larson G, et al. Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. Natl Sci Rev. 2019;6(4):810–824. [PubMed: 31598383]
26. Giannuzzi D, Marconato L, Fanelli A, et al. The genomic landscape of canine diffuse large B-cell lymphoma identifies distinct subtypes with clinical and therapeutic implications. Lab Anim (NY). 2022;51(7):191–202. [PubMed: 35726023]
27. Thomas R, Wiley CA, Droste EL, Robertson J, Inman BA, Breen M. Whole exome sequencing analysis of canine urothelial carcinomas without BRAF V595E mutation: short in-frame deletions in BRAF and MAP2K1 suggest alternative mechanisms for MAPK pathway disruption. PLoS Genet. 2023;19(4):e1010575. [PubMed: 37079639]
28. Amin SB, Anderson KJ, Boudreau CE, et al. Comparative molecular life history of spontaneous canine and human gliomas. Cancer Cell. 2020;37(2):243–257.e247. [PubMed: 32049048]
29. Li H, Durbin R. Fast and accurate long-read alignment with burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–595. [PubMed: 20080505]
30. Li H, Durbin R. Fast and accurate short read alignment with burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–1760. [PubMed: 19451168]

31. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–1111. [PubMed: 19289445]
32. Alsaihati BA, Ho KL, Watson J, et al. Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds. Nat Commun. 2021;12(1):4670. [PubMed: 34344882]
33. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–1303. [PubMed: 20644199]
34. Capriotti E, Montanucci L, Profiti G, et al. Fido-SNP: the first web server for scoring the impact of single nucleotide variants in the dog genome. Nucleic Acids Res. 2019;47(W1):W136–W141. [PubMed: 31114899]
35. Broeckx BJ, Coopman F, Verhoeven GE, et al. Development and performance of a targeted whole exome sequencing enrichment kit for the dog (Canis Familiaris build 3.1). Sci Rep. 2014;4:5597. [PubMed: 24998260]
36. Wong S, Ehrhart EJ, Stewart S, et al. Genomic landscapes of canine splenic angiosarcoma (hemangiosarcoma) contain extensive heterogeneity within and between patients. PloS One. 2022;17(7):e0264986. [PubMed: 35867969]
37. Chu S, Skidmore ZL, Kunisaki J, et al. Unraveling the chaotic genomic landscape of primary and metastatic canine appendicular osteosarcoma with current sequencing technologies and bioinformatic approaches. PloS One. 2021;16(2):e0246443. [PubMed: 33556121]
38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. [PubMed: 20601685]
39. Wilm A, Aw PP, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40(22):11189–11201. [PubMed: 23066108]
40. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213–219. [PubMed: 23396013]
41. Fang LT, Afshar PT, Chhibber A, et al. An ensemble approach to accurately detect somatic mutations using SomaticSeq. Genome Biol. 2015;16(1):197. [PubMed: 26381235]
42. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568–576. [PubMed: 22300766]
43. Zhang M, Liu D, Tang J, et al. SEG–a software program for finding somatic copy number alterations in whole genome sequencing data of cancer. Comput Struct Biotechnol J. 2018;16:335–341. [PubMed: 30258547]
44. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc. 2012;7(3):562–578. [PubMed: 22383036]
45. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–1667. [PubMed: 27560171]
46. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–112. [PubMed: 19847166]
47. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–15550. [PubMed: 16199517]
48. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. [PubMed: 18798982]
49. Liu T. Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol Biol. 2014;1150:81–95. [PubMed: 24743991]
50. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–589. [PubMed: 20513432]

51. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43:D423–D431. [PubMed: 25414341]
52. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. BMC Genom. 2014;15(1):325.
53. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014;30(23):3310–3316. [PubMed: 25143287]
54. Orenbuch R, Filip I, Comito D, Shaman J, Pe'er I, Rabadan R. arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics. 2020;36(1):33–40. [PubMed: 31173059]
55. Kim HJ, Pourmand N. HLA typing from RNA-seq data using hierarchical read weighting [corrected]. PloS One. 2013;8(6):e67885. [PubMed: 23840783]
56. Buchkovich ML, Brown CC, Robasky K, et al. HLAProfiler utilizes k-mer profiles to improve HLA calling accuracy for rare and common alleles in RNA-seq data. Genome Med. 2017;9(1):86. [PubMed: 28954626]
57. Richters MM, Xia H, Campbell KM, Gillanders WE, Griffith OL, Griffith M. Best practices for bioinformatic characterization of neoantigens for clinical utility. Genome Med. 2019;11(1):56. [PubMed: 31462330]
58. Lee H, Kingsford C. Kourami: graph-guided assembly for novel human leukocyte antigen allele discovery. Genome Biol. 2018;19(1):16. [PubMed: 29415772]
59. Bauer DC, Zadoorian A, Wilson LOW, Melbourne Genomics Health A, Thorne NP. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. Brief Bioinform. 2018;19(2):179–187. [PubMed: 27802932]
60. Warren RL, Choe G, Freeman DJ, et al. Derivation of HLA types from shotgun sequence datasets. Genome Med. 2012;4(12):95. [PubMed: 23228053]
61. Boegel S, Lower M, Schafer M, et al. HLA typing from RNA-Seq sequence reads. Genome Med. 2012;4(12):102. [PubMed: 23259685]
62. Miyamae J, Okano M, Nishiya K, et al. Haplotype structures and polymorphisms of dog leukocyte antigen (DLA) class I loci shaped by intralocus and interlocus recombination events. Immunogenetics. 2022;74(2):245–259. [PubMed: 34993565]
63. Miyamae J, Suzuki S, Katakura F, et al. Identification of novel polymorphisms and two distinct haplotype structures in dog leukocyte antigen class I genes: DLA-88, DLA-12 and DLA-64. Immunogenetics. 2018;70(4):237–255. [PubMed: 28951951]
64. Kennedy LJ, Angles JM, Barnes A, et al. Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA nomenclature committee. Tissue Antigens. 2001;58(1):55–70. [PubMed: 11580859]
65. Ross P, Buntzman AS, Vincent BG, et al. Allelic diversity at the DLA-88 locus in Golden retriever and boxer breeds is limited. Tissue Antigens. 2012;80(2):175–183. [PubMed: 22571293]
66. Graumann MB, DeRose SA, Ostrander EA, Storb R. Polymorphism analysis of four canine MHC class I genes. Tissue Antigens. 1998;51(4):374–381. [PubMed: 9583809]
67. Feng Y, Hildebrand WH, Tompkins SM, Zhao S. A Kmer-based paired-end read (KPR) de novo assembler and genotyper to genotype major histocompatibility complex class I (MHC-I) alleles for the dog. bioRxiv. 2007;2020(2020):205559.
68. Dai X, Cai L, He F. Single-cell sequencing: expansion, integration and translation. Brief Funct Genom. 2022;21(4):280–295.
69. Li J, Eberwine J. The successes and future prospects of the linear antisense RNA amplification methodology. Nat Protoc. 2018;13(5):811–818. [PubMed: 29599441]
70. Chen X, Yang Z, Chen W, et al. A multi-center cross-platform single-cell RNA sequencing reference dataset. Sci Data. 2021;8(1):39. [PubMed: 33531477]
71. Ashton JM, Rehrauer H, Myers J, et al. Comparative analysis of single-cell RNA sequencing platforms and methods. J Biomol Tech. 2021;32(4):3fc1f5fe.3eccea01.
72. Chen X, Teichmann SA, Meyer KB. From tissues to cell types and Back: single-cell gene expression analysis of tissue architecture. Ann Rev Biomed Data Sci. 2018;1(1):29–51.

73. Ziegenhain C, Vieth B, Parekh S, et al. Comparative analysis of single-cell RNA sequencing methods. Mol Cell. 2017;65(4):631–643.e634. [PubMed: 28212749]
74. Salomon R, Kaczorowski D, Valdes-Mora F, et al. Droplet-based single cell RNAseq tools: a practical guide. Lab Chip. 2019;19(10):1706–1727. [PubMed: 30997473]
75. Xi NM, Li JJ. Benchmarking computational doublet-detection methods for single-cell RNA sequencing data. Cell Syst. 2021;12(2):176–194.e176. [PubMed: 33338399]
76. Hou W, Ji Z, Ji H, Hicks SC. A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol. 2020;21(1):218. [PubMed: 32854757]
77. Zappia L, Theis FJ. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. Genome Biol. 2021;22(1):301. [PubMed: 34715899]
78. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multi-modal single-cell data. Cell. 2021;184(13):3573–3587.e3529. [PubMed: 34062119]
79. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. [PubMed: 29409532]
80. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15(6):e8746. [PubMed: 31217225]
81. Nayak R, Hasija Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. Genomics. 2021;113(2):606–619. [PubMed: 33485955]
82. Petukhov V, Guo J, Baryawno N, et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. Genome Biol. 2018;19(1):78. [PubMed: 29921301]
83. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017;33(8):1179–1186. [PubMed: 28088763]
84. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst. 2019;8(4):329–337.e324. [PubMed: 30954475]
85. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–172. [PubMed: 30643263]
86. Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. Nat Methods. 2019;16(10):983–986. [PubMed: 31501545]
87. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–386. [PubMed: 24658644]
88. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38(12):1408–1414. [PubMed: 32747759]
89. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford). 2019;2019:1–9.
90. Speir ML, Bhaduri A, Markov NS, et al. UCSC cell browser: visualize your single-cell data. Bioinformatics. 2021;37(23):4578–4580. [PubMed: 34244710]
91. Ayers J, Milner RJ, Cortes-Hinojosa G, et al. Novel application of single-cell next-generation sequencing for determination of intratumoral heterogeneity of canine osteosarcoma cell lines. J Vet Diagn Invest. 2021;33(2):261–278. [PubMed: 33446089]
92. Cambien B, Lebrigand K, Baeri A, et al. Identification of oncolytic vaccinia restriction factors in canine high-grade mammary tumor cells using single-cell transcriptomics. PLoS Pathog. 2020;16(10):e1008660. [PubMed: 33075093]
93. Gingrich AA, Reiter TE, Judge SJ, et al. Comparative immunogenomics of canine natural killer cells as immunotherapy target. Front Immunol. 2021;12:670309. [PubMed: 34594320]
94. Fastres A, Pirottin D, Fievez L, et al. Characterization of the bronchoalveolar lavage fluid by single cell gene expression analysis in healthy dogs: a promising technique. Front Immunol. 2020;11:1707. [PubMed: 32849601]

95. Fastres A, Pirottin D, Fievez L, et al. Identification of pro-fibrotic macrophage populations by single-cell transcriptomic analysis in West Highland White terriers affected with canine idiopathic pulmonary fibrosis. Front Immunol. 2020;11:611749. [PubMed: 33384697]

96. Zhou QJ, Liu X, Zhang L, et al. A single-nucleus transcriptomic atlas of the dog hippocampus reveals the potential relationship between specific cell types and domestication. Natl Sci Rev. 2022;9(11):nwac147. [PubMed: 36569494]

97. Knapp DW, Dhawan D, Ramos-Vara JA, et al. Naturally-occurring invasive urothelial carcinoma in dogs, a unique model to drive advances in managing muscle invasive bladder cancer in humans. Front Oncol. 2019;9:1493. [PubMed: 32039002]

98. Eschke M, Moore PF, Chang H, Alber G, Keller SM. Canine peripheral blood TCRalphabeta T cell atlas: identification of diverse subsets including CD8A(+) MAIT-like cells by combined single-cell transcriptome and V(D)J repertoire analysis. Front Immunol. 2023;14:1123366. [PubMed: 36911660]

99. Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan GC. Advances in spatial transcriptomic data analysis. Genome Res. 2021;31(10):1706–1718. [PubMed: 34599004]

100. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596(7871):211–220. [PubMed: 34381231]

101. Veldman-Jones MH, Brant R, Rooney C, et al. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. Cancer Res. 2015;75(13):2587–2593. [PubMed: 26069246]

102. Garelli CJ, Wong NB, Piedra-Mora C, et al. Shared inflammatory and skin-specific gene signatures reveal common drivers of discoid lupus erythematosus in canines, humans and mice. Curr Res Immunol. 2021;2:41–51. [PubMed: 35492392]

103. Carroll J, Coutermarsh-Ott S, Klahn SL, et al. High intensity focused ultrasound for the treatment of solid tumors: a pilot study in canine cancer patients. Int J Hyperther. 2022;39(1):855–864.

104. Magee K, Marsh IR, Turek MM, et al. Safety and feasibility of an in situ vaccination and immunomodulatory targeted radionuclide combination immuno-radiotherapy approach in a comparative (companion dog) setting. PloS One. 2021;16(8):e0255798. [PubMed: 34383787]

105. Boss MK, Ke Y, Bian L, et al. Therapeutic intervention using a Smad7-based tat protein to treat radiation-induced oral mucositis. Int J Radiat Oncol Biol Phys. 2022;112(3):759–770. [PubMed: 34610386]

106. Polak KZ, Schaffer P, Donaghy D, Zenk MC, Olver CS. Iron, hepcidin, and microcytosis in canine hepatocellular carcinoma. Vet Clin Pathol. 2022;51(2):208–215. [PubMed: 35274348]

107. Harris RA, Rout ED, Yoshimoto JA, Avery PR, Avery AC. Using digital RNA counting to establish flow cytometry diagnostic criteria for subtypes of CD34+ canine acute leukaemia. Vet Comp Oncol. 2022;20(3):710–719. [PubMed: 35491468]

108. Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a cancer dependency map. Cell. 2017;170(3):564–576.e516. [PubMed: 28753430]

109. Maynard LH, Humbert O, Peterson CW, Kiem HP. Genome editing in large animal models. Mol Ther. 2021;29(11):3140–3152. [PubMed: 34601132]

110. Gradauskaite V, Inglebert M, Doench J, et al. LRP6 is a functional receptor for attenuated canine distemper virus. mBio. 2023;14(1):e0311422. [PubMed: 36645301]

111. Das S, Idate R, Cronise KE, Gustafson DL, Duval DL. Identifying candidate druggable targets in canine cancer cell lines using whole-exome sequencing. Mol Cancer Ther. 2019;18(8):1460–1471. [PubMed: 31175136]

112. Goodsaid FM. The labyrinth of product development and regulatory approvals in liquid biopsy diagnostics. Clin Transl Sci. 2019;12(5):431–439. [PubMed: 31162800]

113. de Rubis G, Rajeev Krishnan S, Bebawy M. Liquid biopsies in cancer diagnosis, monitoring, and prognosis. Trends Pharmacol Sci. 2019;40(3):172–186. [PubMed: 30736982]

114. Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun. 2017;8(1):1324. [PubMed: 29109393]

115. Garcia-Pardo M, Makarem M, Li JJN, Kelly D, Leighl NB. Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges. Br J Cancer. 2022;127(4):592–602. [PubMed: 35347327]

116. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic–implementation issues and future challenges. Nat Rev Clin Oncol. 2021;18(5):297–312. [PubMed: 33473219]

117. Mochizuki H, Kennedy K, Shapiro SG, Breen M. BRAF mutations in canine cancers. PloS One. 2015;10(6):e0129534. [PubMed: 26053201]

118. Decker B, Parker HG, Dhawan D, et al. Homologous mutation to human BRAF V600E is common in naturally occurring canine bladder cancer–evidence for a relevant model system and urine-based diagnostic test. Mol Cancer Res. 2015;13(6):993–1002. [PubMed: 25767210]

119. Tagawa M, Tambo N, Maezawa M, et al. Quantitative analysis of the BRAF V595E mutation in plasma cell-free DNA from dogs with urothelial carcinoma. PloS One. 2020;15(4):e0232365. [PubMed: 32330187]

120. Flory A, Kruglyak KM, Tynan JA, et al. Clinical validation of a next-generation sequencing-based multi-cancer early detection “liquid biopsy” blood test in over 1000 dogs using an independent testing set: the CANcer detection in dogs (CANDiD) study. PloS One. 2022;17(4):e0266623. [PubMed: 35471999]

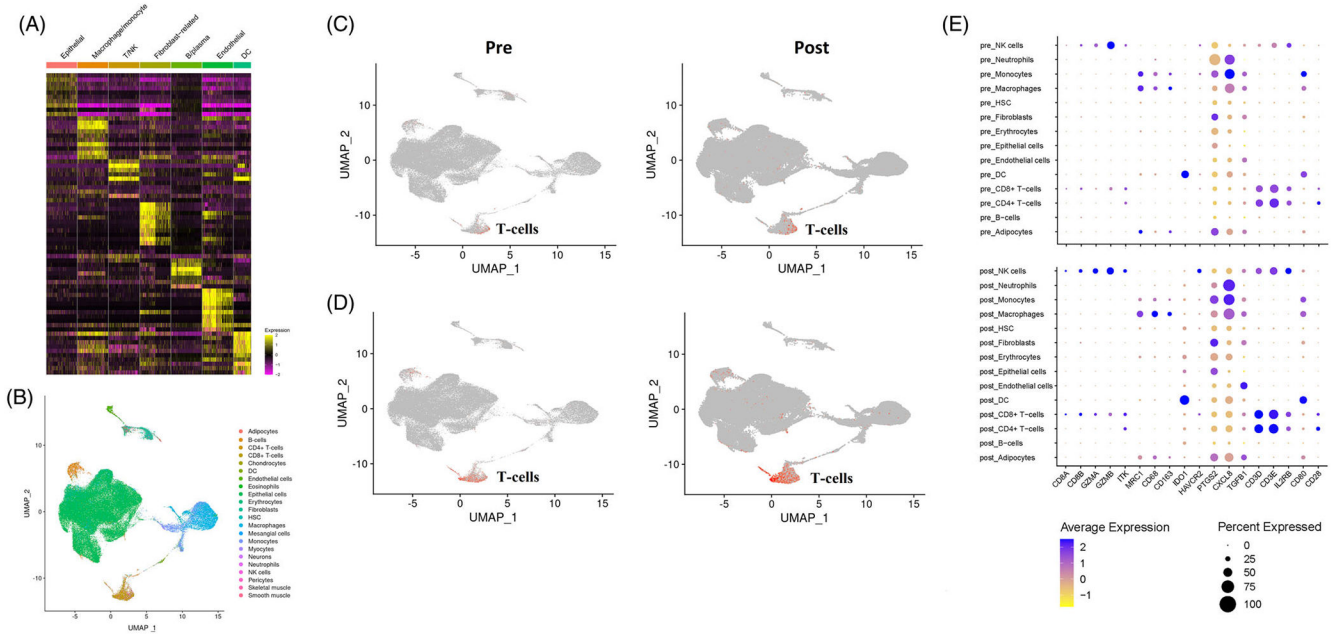

FIGURE 1.

Analysis of scRNA-seq data from canine invasive urothelial carcinoma. This analysis was part of a clinical study in dogs with naturally occurring invasive bladder cancer aimed at defining mechanisms responsible for the anti-tumour effects of cyclooxygenase inhibiting drugs. Briefly, invasive urothelial carcinoma tissues were collected by cystoscopic biopsy before and 1 month following the initiation of cyclooxygenase inhibitor treatment. The tumour samples were digested with collagenase and filtered. Cells were sorted by flow-activated cell sorting and used to generate gel bead-in emulsions (GEMs) using the 10× Chromium system following manufacturer's protocol (10× Genomics, Pleasanton, CA). The in-GEM reverse transcription (RT) was followed by cDNA clean-up and quality control. Libraries were prepared using 10× Chromium Single Cell 3′ Reagents Kits and submitted for sequencing (Novogene, Sacramento, CA). The results demonstrated an increase in the expression of CD8 and other genes, and the cell types involved. (A) Heatmap representing gene expression in specific cell types.

(B) Representative plot with cell type assignment. (C) and (D) Plots of gene expression changes (C: CD8B, D: CD3D) in biopsies collected pre and post 4 weeks cyclooxygenase inhibitor treatment. (E) Dot plot summarizing changes in expression of multiple immune-related genes in specific cell types pre- (top panel) and post cyclooxygenase inhibitor (lower panel). Figure courtesy of D. Dhawan and S. Utturkar.

TABLE 1

Available canine genomes.

GenBank Accession ID	Year	Source	Name	Breed	Genome	Alias
GCA_000002285.1	July 2004	Broad & Agencourt Bioscience		Boxer	canFam1	Out of Date
GCA_000002285.2	May 2005	Broad & Beckman Coulter		Boxer	canFam2	Out of Date
GCA_000002285.3	September 2011	Broad & Beckman Coulter Genomics		Boxer	canFam3	Broad CanFam 3.1
GCA_011100685.1	March 2020	Uppsala University	Mischka	German Shepherd	canFam4	UU_Cfam_GSD_1.0
GCA_005444595.1	May 2019	University of Michigan	Zoey	Great Dane	canFam5	UMICH_Zoey_3.1
GCF_000002285.5	October 2020	Dog Genome Sequencing Consortium	Tasha	Boxer	canFam6	Dog10k_Boxer_Tasha
GCF_014441545.1	September 2020	The Roslin Institute		Labrador retriever		ROS_Cfam_1.0

TABLE 2

Analytic tools.

Tool	Function	Source and documentation
Trimmomatic (version 0.39)	Sequence trimming by removing adapter sequences, low quality bases, et.	http://www.usadellab.org/cms/?page=trimmomatic
FastQC	Sequence quality control	https://github.com/s-andrews/FastQC
ANNOVAR[38]	Functional annotation of sequencing variants	https://doc-openbio.readthedocs.io/projects/annovar/en/latest/
Funcotator	Functional annotation of sequencing variants	https://gatk.broadinstitute.org/hc/en-us/articles/360037224432-Funcotator
BWA[29,30] (v. 07.17)	DNA sequence alignment	http://bio-bwa.sourceforge.net/
Genome Analysis ToolKit (GATK)[33] (v. 3.8.1)	DNA sequence alignment and analysis	https://software.broadinstitute.org/gatk/
LoFreq[39] (v. 2.1.2)	DNA variant caller	https://github.com/CSB5/lofreq
Mutect2–GATK4[40] (v. 4.1.6)	DNA variant caller	https://software.broadinstitute.org/gatk/
Picard (v. 2.16.0)	NGS data manipulator	http://broadinstitute.github.io/picard
Somatic Seq[41] (v. 3.4.1)	DNA variant caller	https://github.com/bioinform/somaticseq
VarScan2[42] (v. 2.4.2)	DNA variant caller	https://github.com/dkoboldt/varscan
Strelka2	DNA variant caller	https://github.com/Illumina/strelka
Manta	DNA variant caller	https://github.com/Illumina/manta
SignatureAnalyzer[32]	Mutation signatures	https://software.broadinstitute.org/cancer/cga/msp
DNAcopy (v. 1.62.0)	DNA copy number caller	https://github.com/veseshan/DNAcopy
SEG[43]	DNA copy number caller	https://github.com/ZhaoS-Lab/SEG
Tophat v2.0.14[31,44]	RNA aligner	http://ccb.jhu.edu/software/tophat/index.shtml
HISAT v2.1.0[45]	RNA aligner	https://ccb.jhu.edu/software/hisat2/index.shtml
Cufflinks v2.2.1[44]	Expression analysis	http://cole-trapnell-lab.github.io/cufflinks
ssGSEA v9.0.10[46]	Gene set enrichment analysis	http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/ssGSEAProjection/4
GSEA v3.0[47]	Gene set enrichment analysis	https://github.com/GSEA-MSigDB/gsea-desktop
MACS v2.1.1[48,49]	ChIP-Seq analysis	https://github.com/taoliu/MACS
HOMER v4.10[50]	Motif analysis	http://homer.ucsd.edu/homer/

TABLE 3

Examples of commonly utilized bioinformatic tools for scRNA-seq.

Category	Tool	Function	Source and documentation
Pre-processing	CellRanger	Analysis pipeline that process chromium (10× genomics) single cell data to align reads, generate feature-barcode matrices, perform clustering and other secondary analysis	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
	dropEst[82]	Analysis pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments	https://dropest.readthedocs.io/en/latest/
Integrated workflow	Seurat[78]	R-package designed to provide-to-workflow (quality control, analysis, visualization and integration) of single-cell data	https://satijalab.org/seurat/index.html
	Scanpy[79]	A scalable toolkit designed to provide-to-workflow (pre-processing, visualization, clustering, trajectory inference and differential expression) of single-cell RNAseq data	https://scanpy.readthedocs.io/en/stable/
Quality control	scater[83]	A collection of tools for doing various analyses of single-cell RNA-seq gene expression data, with a focus on quality control and visualization	https://bioconductor.org/packages/release/bioc/html/scater.html
	DoubletFinder[84]	Doublet detection in single-cell RNA sequencing data using artificial nearest neighbours	https://github.com/chris-mcginnis-ucsf/DoubletFinder
Cell type identification	SingleR[85]	Unbiased cell type recognition from scRNAseq data using reference transcriptomic datasets of pure cell types	https://bioconductor.org/packages/release/bioc/html/SingleR.html
	Garnett[86]	Automated cell type classification from scRNAseq data using supervised classification	https://cole-trapnell-lab.github.io/garnett/docs/
Pseudotime-trajectory analysis	Monocle[87]	Constructing single-cell trajectories (pseudotime) and differential expression analysis	https://cole-trapnell-lab.github.io/monocle3/
	scVelo[88]	A scalable toolkit for RNA velocity analysis in single cells through dynamical modelling	https://scvelo.readthedocs.io/
scRNAseq repositories	PanglaoDB[89]	Unified database for exploration of single cell RNA sequencing experiments from mouse and human	https://panglaodb.se/
	UCSC Cell Browser[90]	Interactive viewer for single-cell data	https://cellbrowser.readthedocs.io/en/master/

Note: Tools defined under ‘pre-processing’, ‘integrated-workflow’, and ‘quality-control’ are applicable to canine data; these are generic tools and can be applied as long as references and annotations are provided. The tools under ‘cell-type identification’ can be applied to canine data with orthologous gene-symbols. The human-based classifiers are used with the orthologous gene-symbols. These assignments are treated as ‘guide only’, and manual curation with canine genes is applied for final cell type assignment. The ‘pseudotime-trajectory analysis’ tools can be performed using reference/annotation, although application of these tools for canine data has not been reported. The scRNA-seq repositories are dominated by human/mouse studies along with a small number of studies in other species. This is an area where the Integrated Canine Data Commons can contribute canine scRNA-seq data in the future.