Table of Contents

Max 6 sider

Introduction

In recent years, the domestic dog, Canis lupus familiaris, has become an increasingly useful comparative spontaneous cancer model to study genetic and environmental risk factors as well as easing the transition between rodent and human clinical trials for cancer drug development. The domestic dog has become. The many similarities between various cancer types affecting humans and dogs and the spontaneous development of these cancers in immune-competent canine individuals living in a shared environment with us suggest a common etiology. The shorter lifespan of dogs and the shorter time to relapse after cancer treatment allows data regarding efficacy, short- and long-term toxicity and side effects of novel cancer drugs to be generated in years rather than decades as in human clinical trials. However, certain limitations need to be overcome to make full use of the dog model. Different classification systems for common cancers limit translation of data and clinical outcome from dog to human. The current canine genome and annotation impairs a careful and correct comparison with the human genome.

While one of the biggest advantages of dogs is the breed structure which makes mapping predisposing risk factors relatively easy, the fact that dogs also share the human environment and receive high-quality health care makes them even better models for cancer.

We will focus on three complex cancers: breast cancer, lymphoma, and osteosarcoma. where improved models would benefit human studies, and the canine forms are diverse and in need of a more comprehensive characterization. Since the molecular sub-classifications used in human cancers are not used for these canine cancers, we plan to characterize these canine tumor types molecularly to meet human standards.

Aim

Define comparable molecular phenotypes for lymphoma, osteosarcoma, and breast cancer and collect longitudinal samples for mammary tumor evolution analysis**.**

Developing the dog as a model for human cancer remains an underutilized resource. We will use existing T/N WES to compare breeds and histopathological subtypes for lymphoma, osteosarcoma, and breast cancer. We will use this to select and prioritize existing and prospective T/N samples for RNA sequencing as well as tumor and normal WGS. We will use this to identify comparable molecular subgroups for all three cancers. We will select dogs with malignant mammary tumors and resample plasma every month for cfDNA sequencing as a proxy for tumor mutations [18]. This will allow us to study tumor evolution and develop infrastructure necessary for performing clinical trials in the future.

Breast cancer studies through mammary tumors

In canines, mammary tumors are the most common type of neoplasia affecting non-neutered female dogs [2]. The tumors can be divided into subgroups mainly based on their histopathological features. Attempts to classify canine mammary tumors using similar classification system as in humans have not been successful as they have been based on immunohistochemistry using poorly validated antibodies or on expression data for a small panel of genes and individuals [41-44]. The utility of canine mammary tumors as a model for human breast cancer in clinical trials has therefore been limited. Still, the route to metastasis for canine mammary tumors is similar to human breast cancer. Infiltration of the local lymph node or distant metastasis to lung tissue, bone or brain makes the dog

a good model for metastatic disease [41, 45]. Hence, dogs would provide a good model for human metastatic breast cancer if subtyping challenges could be overcome using molecular tools.

Lymphoma, a heterogeneous cancer in both species.

Among both humans and dogs, diffuse large B-cell lymphomas (DLBCL) make up the majority of non-Hodgkin's lymphoma cases. Human DLBCL is classified into ABC and GCB subtypes based on gene expression. While similar subtypes have been suggested for canine DLBCL [1], additional studies are needed to confirm which canine lymphomas are good models for specific human subtypes. In addition, peripheral T-cell lymphoma not otherwise specified (PTCL-NOS) is relatively rare in humans and has a poor prognosis. PTCL-NOS is considerably more common among dogs [3], for example in the boxer breed. Since murine transgenic or knock-out PTCL-NOS models vary in how well they recapitulate the human phenotype [4], a canine model with spontaneously occurring disease could greatly improve our understanding. Detailed matching of subtypes between the species will greatly inform the translation of canine clinical data to human research. The social and economic impact of lymphoma is high, both for dogs and humans.

Osteosarcoma, common in large dog breeds.

Human osteosarcoma is most common in teenagers and young adults. It is a relatively rare, but often deadly disease. In large dog breeds osteosarcoma is more prevalent than in other breeds, with close to 13% of Rottweilers getting osteosarcoma. Like in humans [5], male dogs are slightly more likely to acquire the disease compared to females [6], allowing evaluation of hormone-related factors. Human studies have pointed to a high rate of TP53 mutations and copy number aberrations in human tumors. Still, human studies are limited by the rarity of the disease, and a good model system is needed.

Innovation

For the first time, we will use state-of the art tumor/normal sequencing and expression profiling to generate molecular profiles and develop gene panels for identifying which canine patients match which specific human subtypes of breast cancer, lymphoma, and osteosarcoma, respectively. In addition, we will enrich these expression profiles by comparing them with canine germ-line risk factors and tumor mutations found in the corresponding dogs. We will further enhance the comparative value of the dog model by including both purebred dogs and regular mixed breed family dogs. We will work closely with human and canine cancer experts on breast cancer, lymphoma and osteosarcoma to ensure state of the art matching of molecular phenotypes.

Background And Preliminary Results

Genome-wide association mapping and functional analysis of canine mammary tumors. Using the Illumina 170K canine HD SNP array [7], we have performed a genome-wide association study (GWAS) in 332 ESS (188 cases, 144 controls) [8]. EMMAX software [9] was used to calculate association p-values corrected for stratification and cryptic relatedness using mixed model statistics. Eight SNPs, located in three genomic regions, show genome-wide significant p-values. These three top associated loci explain 28.1±10.0% of the phenotypic variation. In addition to the associated regions, an almost completely fixed region was identified on chromosome 30. Following resequencing and further fine-mapping, multiple associated haplotypes on chromosome 11 spanning the CDK5RAP2 gene and parts of the MEGF9 gene and a small gene desert between CDK5RAP2 and DBC1 were identified. The most strongly associated SNP is located within a small gene desert

upstream of the CDK5RAP2 gene. The risk allele disrupts a transcription factor binding motif for PNR/NR2E3, which is an orphan nuclear hormone receptor. It is a regulator of the estrogen receptor 1 (ESR1) in ER positive breast cancer cells and interleukin 13Ra2 in ER negative breast cancer cells, regulating tumor growth, cell migration and metastasis [10, 40].

In addition to the published GWA for ESS [8] we are in the process of analyzing GWA datasets from additional breeds including cocker spaniels, German shepherds and boxers as well as additional ESS.

Genome-wide association mapping and functional analysis of lymphoma. Using the Illumina 170K canine HD SNP array, we have performed a GWAS of B-cell lymphoma cases and controls among Golden retrievers identifying two predisposing loci that are shared with hemangiosarcoma [11]. Investigation of gene expression changes in B-cell lymphomas from individuals with risk or non-risk haplotypes identified significant changes in genes associated with immune cell function. In addition, a GWAS in the same breed studying T-zone lymphoma cases and controls identified two predisposing loci that are not shared with other T-cell lymphoma subtypes (unpublished data with collaborator Anne Avery, Colorado State University).

Genome-wide association mapping and functional analysis of osteosarcoma. Using the same Illumina 170K canine HD SNP array as above we investigated the genetic predisposition to osteosarcoma in three different breeds. We identified 33 associated loci, most of which were specific to the different breeds, but implying genes in shared pathways [2]. Detailed analysis of the top Greyhound locus near CDKN2A/B identified the causative SNP and showed that it alters a transcription factor binding site. The risk variant was fixed in the two other breeds investigated. Five additional breeds are currently in GWAS genotyping.

T/N analysis of lymphoma. Using an exome array developed by my group, we set out to study tumor mutations in canine B- and T-cell lymphoma in three breeds with differential predisposition [12]. In the two breeds with B-cell lymphoma, we saw strong similarities between golden retrievers and cocker spaniels, with typical mutations in POT1 (17% of all cases), FBXW7 (25%), and TRAF3 and/or MAP3K14 (28%). 40% of the FBXW7 mutations occur in a specific codon; the corresponding codon is recurrently mutated in human cancer. In contrast, little overlap was seen in the mutated genes in the two T-cell lymphoma predisposed breeds, boxers and golden retrievers. They only share one of their 27 significantly mutated genes. Boxers, who develop aggressive T-cell lymphomas, are typically mutated in the PTEN-mTOR pathway. T-cell lymphomas in golden retrievers are often less aggressive and the studied tumors typically show mutations in genes involved in cellular metabolism. Some overlap was seen between the significantly mutated genes in golden retrievers with B- and T-cell lymphoma. We identified genes with known involvement in human lymphoma and leukemia, genes implicated in other human cancers, as well as novel genes that could allow new therapeutic targets.

T/N analysis of osteosarcoma. A similar analysis of T/N WES data from three breeds with osteosarcoma is currently ongoing. Preliminary data point to strong similarities with human osteosarcoma. The most frequently mutated gene, TP53, showed mutations in ~60% of all dogs. Copy number detection by VarScan2 [13] and GISTIC2 [14] showed extensive somatic copy number aberrations (SCNAs) across all three breeds. Both TP53 mutations and large numbers of SCNAs are hallmarks of human osteosarcoma. In addition, as a novel OSA finding, SETD2 was mutated in ~21% of dog samples (golden retriever 32% (n=22), Rottweiler 19% (n=21) and greyhound 13% (n=23)) and showed a variety of mutation types (frameshift, nonsense, splice and missense). The SETD2 gene encodes a histone methyltransferase and has been implicated as a tumor suppressor gene in a number of cancers [15, 16]. Intriguingly, analysis of mutational signatures [17] identified three signatures at

different frequencies across breeds. They included Cosmic 1, often related to aging, Cosmic 17 and a novel signature seen more frequently in tumors from golden retrievers.

Experimental Approach

Sample collection strategies and sample preparation:

We have sequenced and analyzed T/N WES data for B-cell (2 breeds) and T-cell (2 breeds) lymphoma and osteosarcoma (3 breeds), and T/N WES for mammary tumors is under way (3 breeds + additional samples). B-cell lymphomas from 2 breeds are remarkably similar whereas T-cell lymphomas from 2 breeds exhibit mutations in unrelated pathways [12]. For osteosarcoma, the same pathways are mutated across breeds, but mutational load correlates with breed predisposition (manuscript submitted). We will perform RNA-Seq and WGS to evaluate how molecular subtypes correlate with human cancer subtypes. Lymphoma samples will be selected based on histological subtype and breed. Osteosarcoma samples with known recurrence outcome from three breeds will be prioritized to allow for analysis of expression patterns in different genetic backgrounds. Mammary tumor WES (underway) will be analyzed before new samples are selected for sequencing. Veterinary oncologists and pathologists will be consulted for all samples. For many breeds in each cancer we have performed GWAS which allows correlation of germ-line and somatic variants.

Throughout the life of the project – using the newly formed consortium described below and other collaborators – we will collect tumor and matched normal tissue from all dogs with mammary tumors, all major lymphoma subtypes, and osteosarcomas with known outcome. We currently have around 80 mammary tumors and 30 lymphomas and we have access to 200 osteosarcomas (many from CCOGC). We will sequence (WGS and RNA-Seq) and analyze ~40 T/N pairs each year, focusing on one of the three target cancers each year. At least 10 samples will be sequenced for each breed/subtype as this number is sufficient to identify significantly mutated genes [12]. We will evaluate gene expression patterns, analyze the mutational spectrum in dog breeds predisposed to multiple cancers, and correlate significantly mutated genes with clinical outcome. We will extract both DNA and RNA from RNA-later preserved tumor tissue using the All-Prep RNA/DNA extraction kit. DNA and RNA quality will be evaluated using the Agilent tape station. We will use DNA extracted from peripheral blood to represent germ-line variance for each dog.

We will also collect plasma for monthly sampling of select dogs with mammary tumors. These dogs will be recruited during year 1-2 and sampled for 18 months for cell-free DNA sequencing. Samples will be collected in Streck Cell-Free DNA blood collection tubes.

RNA-Seq for expression analysis: RNA samples from tumor tissue collected as described above from the key breeds and a mix of breeds, will be sent for polyA selected strand specific RNA-Seq library preparation at the SciLifeLab Sequencing Platform and subsequent sequencing on Illumina HiSeq 2500 to allow for at least 50x coverage of the transcriptome. MicroRNAs and lincRNAs will also be sequenced. The RNA sequencing data will be mapped to the canine reference genome CanFam4.0 using STAR [19]. Differential expression between tumor types will be assessed using CuffDiff and cummeRbund [20] and edgeR [21]. Genes differentially expressed between tumors with a specific histopathological diagnosis or between tumors from different breeds will be confirmed using rtPCR on cDNA. Pathway analysis will be performed, using DAVID [22] and PathwAX [23] to detect pathways that are alternatively expressed in different tumors. Both breed-based and histology-based comparisons will be performed.

Tumor/normal analysis: T/N WGS data will be mapped to the canine reference genome CanFam4.0 using the Burrows-Wheeler Aligner [24], and duplicates will be marked using Picard (broadinstitute.github.io/picard/). Subsequent base quality recalibration and indel realignment will be performed with the genome analysis toolkit (GATK, [25]). Finally, variant discovery will be performed using muTect2 [26, 27] inputting a panel of normal variants previously detected across breeds and cancers ([12]), and significantly mutated genes will be identified using Genome MuSiC [28]. Copy number aberrations will be identified using VarScan2 [29] and analyzed using GISTIC2 [14]. Other structural variation in tumor and normal genomes will be assessed using a combination of tools using different types of information (split reads, discordant read pairs, read count) and specialized for detecting genomic rearrangement of different sizes: Lumpy [30], Delly [31], Pindel [32, 33], and SVDetect [34]. All have successfully been used to analyze cancer genomes. Similarities and differences of mutated genes, copy number aberrations, and other SVs, and specific mutations in key genes will be compared among all the individuals as well as between breeds. The catalog of detected canine tumor mutations will be compared to tumor mutations identified in human breast cancer studies [35]. For each dog we will also compare germ-line and somatic mutations and copy number aberrations. Here utilizing the CanFam4.0 and the generated catalogs of germ-line variation including CNVs will be important.

Detection of mutational signatures: We will also examine the mutational signature landscape for the somatic mutations identified across the tumor genome. For mutational signatures, we use the Bayesian non-negative matrix factorization (NMF) methodology developed at the Broad Institute [36], which we have successfully applied to WES of 66 canine OSA T/N pairs. For other canine cancers, we have found several of the signatures observed in human cancers (i.e. Cosmic 1 related to aging) [17, 37], but also a novel signature that we are currently exploring. We have detected both breed-specific and cancer-specific signatures (Ingegerd Elvers unpublished data).

Dissection of molecular phenotypes: We will evaluate molecular phenotypes by performing unsupervised clustering of differentially expressed genes across tumors from the same cancer and determining if there are any recurrent patterns. Using significantly mutated genes and SCNAs and other SVs from T/N WGS, we will evaluate whether the identified subgroups are associated with breeds or location/histological subtype or mutations in specific pathways or mutations in recurrent combinations of genes. This analysis will be performed at the beginning of the project with the current T/N WES datasets, and later on with the T/N WGS data generated as part of this project.

Comparison to human cancer: We plan to compare the molecular profiles identified in each of our tumor types with existing human data, thereby evaluating which molecular profiles the canine models represent. Significantly mutated genes, and their mutations as well as affected pathways will be compared with existing human data for each tumor type. We will utilize data from The Cancer Genome Atlas (TCGA, https://tcgadata.nci.nih.gov/docs/publications/tcga/, now served at Genomic Data Commons: https://gdc.cancer.gov) which reports on >1000 breast cancer patients, ~ 380 osteosarcoma patients. In addition, we will use data from the COSMIC database (containing data from thousands of tumors of each type). We will perform clustering to see if similar patient subgroups are shared between the two species. Based on existing canine tumor expression data we expect overlaps with human subtypes of the same cancer, but we will also look for canine models for different human cancers (similar to how canine TCC commonly display BRAF mutations, making them a better model for human BRAF mutated cancers than human TCC [38]. Such models are relevant since for example many but not all BRAF mutated cancers respond to the same drugs [39]). We will look for new and existing therapeutic targets within our datasets to evaluate the suitability for each tumor subtype for clinical trials. We will share our canine cancer data both on the UCSC

Track Hub (where specific mutations, affected gene and eQTLs can be visualized) as well as in the NCIP forum.

The tools and platforms generated through the COHA efforts led by Dr. London will be leveraged to facilitate the development of the proposed trials consortium in Europe and permit synchronization with ongoing efforts in the US, the ultimate goal of creating a coordinated multi-national network driving comparative and translational medicine.

Longitudinal cfDNA sampling to examine tumor evolution: We will select 10 malignant mammary tumors and perform repeat cell-free DNA sampling every month for 12 months after the dogs' surgery, utilizing their faster progression compared to human breast cancer (LOS Adalsteinsson). These samples will be subjected to 0.1x whole genome coverage (ULP-WGS) for detection of the cell free tumor DNA fraction based on the CNAs detected using ichorCNA [18]. If tumor DNA constitutes >10% of total cfDNA the samples will be submitted to WES to detect coding tumor mutations. cfDNA has been previously successfully studied in canine mammary tumors [87]. A whole-body CT scan will be performed after 9 months. This pilot dataset will allow us to ask several questions: Can we detect cell free DNA in plasma from mammary tumors? Are the amounts of cfDNA stable over time? What is the progressive evolution of tumor mutations in these malignant samples? Is there a correlation with outcome?

Potential pitfalls and alternative solutions:

Sample collection. The majority of samples have already been collected, using standardized reagents to ensure RNA preservation. Great care will be taken in extracting RNA from tissues to preserve RNA quality, and this factor will be monitored throughout the process to remove low-quality samples. The majority of previously collected canine samples lack information on treatment and survival, so this aspect cannot be matched to human data. However, as we collect samples we will collect clinical follow-up going forward.

Molecular subtypes. RNA mutations will be validated using WGS data. Multiple software will be used and compared for pathway analysis (IPA, DAVID). cfDNA sequencing. If cfDNA levels are insufficient during the first 3 months in the first 3 dogs, focus will be shifted to metastatic patients.

Conclusion

We plan to improve the dog as a cancer model by specifically improving the use of mammary tumors as a model for human breast cancer, and. We will evaluate different molecular and genetic aspects of canine mammary tumors to allow the classification into comparative human subtypes. The goal is to allow for better translational value of this model in clinical trials. We aim to take advantage of the relatively homologous genome structure within individual dog breeds to investigate the connection between RNA expression, somatic mutations and germ-line risk factors, a feature which is difficult in humans due to their more diverse genetic variation. To support this work as well as the translational studies of various other canine cancers, we plan to organize a Scandinavian Veterinary Oncology Consortium. Through this, we will assemble a platform to share research findings quickly, to enable effective cancer sample collection as well as to have an organization for clinical trials in the future. We hypothesize that this information can be of great value for finding targets for immunotherapy, for understanding cell-of-origin in lymphomas, and for evaluating genetic risk factors predisposing to cancer.