Integrated Canine Data Commons
Data Governance Advisory Board
Prototype Submission

1. Name/Identifier of Study
Zhaolab-CAN-14-0392

2. Grant ID and funding source (if applicable)
NCI R01 CA182093

3. IACUC/IRB approval numbers (if applicable)
A2019 12-001-Y1-A0

4. Scientific Point of Contact (Name, Phone, Email)
Dr. Shaying Zhao
706-542-0145
szhao@uga.edu

5. Data Manager Point of Contact (Name, Phone, Email)
Dr. Shaying Zhao
706-542-0145
szhao@uga.edu

Joshua Watson
706-583-0854
jlw87619@uga.edu

6. Data access policy (choose one): Open-access – no-embargo, Controlled-access – no embargo, Open-access – embargo, Controlled-access - embargo
Open-access – no-embargo
Sequence data have been submitted to the NCBI SRA database with accession numbers SRP023115, SRP023472 and SRP024250. aCGH data have been submitted to the GEO database with accession number GSE54535.

7. Cancer type(s) included in study
Canine mammary cancer including the following subtypes
-Simple carcinoma
-Complex carcinoma & adenoma

8. Number of subjects included in study
16

9. Data types included in study (check all that apply): Imaging, genomics, proteomics, immunology, clinical, other (specify)
Whole genome sequencing (WGS)
Whole exome sequencing (WES)

RNA-seq
Array comparative genome hybridization (aCGH)

10. Amount of data (in TB)
~ 0.33 TB estimated for raw compressed fastq sequencing data.
~ 0.67 TB estimated for mapping (bam files), other processed files and results (mutation, CNA, fusion genes, expression, etc.)
~ 1 TB total

11. The overall scientific benefit of including this study in the ICDC prototype.
Our study contains a comprehensive, high-quality, manageable amount genomic sequencing data of canine mammary cancer, providing an ideal dataset for a prototype submission to the ICDC. Our data will be useful for troubleshooting the pipelines for future genomic data submission to the ICDC by other groups.

The study is already published at https://cancerres.aacrjournals.org/content/74/18/5045.long. The title and abstract is pated below.

Molecular Homology and Difference between Spontaneous Canine Mammary Cancer and Human Breast Cancer

Abstract

Spontaneously occurring canine mammary cancer represents an excellent model of human breast cancer, but is greatly understudied. To better use this valuable resource, we performed whole-genome sequencing, whole-exome sequencing, RNA-seq, and/or high-density arrays on twelve canine mammary cancer cases, including seven simple carcinomas and four complex carcinomas. Canine simple carcinomas, which histologically match human breast carcinomas, harbor extensive genomic aberrations, many of which faithfully recapitulate key features of human breast cancer. Canine complex carcinomas, which are characterized by proliferation of both luminal and myoepithelial cells and are rare in human breast cancer, seem to lack genomic abnormalities. Instead, these tumors have about 35 chromatin-modification genes downregulated and are abnormally enriched with active histone modification H4-acetylation, whereas aberrantly depleted with repressive histone modification H3K9me3. Our findings indicate the likelihood that canine simple carcinomas arise from genomic aberrations, whereas complex carcinomas originate from epigenomic alterations, reinforcing their unique value. Canine complex carcinomas offer an ideal system to study myoepithelial cells, the second major cell lineage of the mammary gland. Canine simple carcinomas, which faithfully represent human breast carcinomas at the molecular level, provide indispensable models for basic and translational breast cancer research.

12. Any publications associated with this study, if any.
Liu, D., et al. (2014). "Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer." Cancer research **74**(18): 5045-5056.

13. Time constraints on processing/loading/releasing the data
Our data are freely available; so no time restraints are anticipated.

14. Data standards used, if any (e.g., SEND)
NA

15. Anticipated budget needed to prepare data set for submission.
Our data are already published and released; hence no additional budget is anticipated for the submission.