RNASEQ ANALYSIS OF CANINE COLORECTAL CANCER AND FUNCTIONAL STUDY OF C1ORF63

by

YE WANG

(Under the Direction of Shaying Zhao)

ABSTRACT

In this study, by analyzing the RNAseq data of 25 canine colon tumor and normal samples, we want to further explore the potentiality of canine colorectal cancer as a model of human colorectal cancer. The colon cancer samples were separated into 3 groups based on PCA analysis. By integrating the histology data and PCA analysis results, the three groups are two different groups of carcinomas and one normal group. The DEseq analysis shows that, one of the carcinoma groups has highly regulated cell cycle signaling and the other carcinoma group has highly regulated immune-response signaling. They also have different residential bacteria. Future work can be done to analyze genetic alterations of these two groups and then compare to the human study of hypermutated and non-hypermutated carcinomas. A gene named C1orf63 which is related to cell proliferation is also studied.

INDEX WORDS: Canine Colorectal Cancer, RNAseq, C1orf63, Cell Proliferation

RNASEQ ANALYSIS OF CANINE COLORECTAL CANCER AND FUNCTIONAL STUDY OF C1ORF63

by

YE WANG

MS, Department of Biochemistry and Molecular Biology, Third Year

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2016

© 2016

Ye Wang

All Rights Reserved

RNASEQ ANALYSIS OF CANINE COLORECTAL CANCER AND FUNCTIONAL STUDY

OF C1ORF63

by

YE WANG

Major Professor:	Shaying Zhao
Committee:	Jianfu Chen
	Kevin K Dobbin

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2016

TABLE OF CONTENTS

Page

LIST OF FIGURES v

CHAPTER

1 Introduction 1

2 Materials and Methods 5

Tophat mapping 5

Cufflinks and Cuffdiff 5

PCA and DEseq 6

Inflammation study 6

Functional study of C1orf63 7

3 Results 8

PCA analysis reveals three groups among all the sequenced samples 8

DEseq analysis shows that one carcinoma group has highly regulated cell cycle signaling and the other carcinoma group has highly regulated immune-response signaling 11

Inflammation study shows that bacteria in the two carcinoma groups are different 13

C1orf63 expression was increased in Taxol treated cells 15

4 Discussion 17

REFERENCES 19

LIST OF FIGURES

Page

Figure 1: PCA analysis with all the genes that have at least 1 FPKM >=1 8

Figure 2: PCA analysis of CRC samples with metabolic gene FPKM values 10

Figure 3: DEseq analysis heatmap 12

Figure 4: Pathogen detection with RNAseq data 14

Figure 5: C1orf63 expression level in Taxol treated HCT116 15

Figure 6: Immunofluorescence staining of C1orf63 16

INTRODUCTION

Colorectal cancer is the third most common cancer. It is also the second leading cause of cancer death in the United States, second to lung cancer. TCGA(The Cancer Genome Atlas) has published a study of human colon cancer named "Comprehensive molecular characterization of human colon and rectal cancer".[1] In this paper, they thoroughly characterized the somatic alterations of human colon and rectal cancer by studying the exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression data. They found that 16% of colorectal carcinomas were hypermutated. The causation of hypermutation in colorectal cancer includes microsatellite instability, somatic mismatch-repair gene and polymerase E mutations. The other colorectal cancers beside the hypermutated colorectal cancers share similar genetic alterations, including gene mutations, copy-number alterations and chromosomal translocations. These findings will potentially help the drug target design and marker detection for clinical therapeutics of human colorectal cancer.

Dog as the intimate company of human has been shown to be a great model organism in the study of certain human cancers as has been showed in the following papers.[2][3] Dog as a model in cancer study has its innate advantages. Pet dogs share similar surroundings as human, which determines that they are exposed to similar carcinogens.[3] Especially as to colorectal cancer, dog sometimes shares human food, which is closely related to initiation of colorectal cancer. In addition, the sequencing depth of dog genome is high enough to serve as a model organism.[4] Besides, dog has spontaneous accruing cancer, not like mouse. Based on the above advantages, using dog as a model organism will avoid artifacts during study and largely reveal

the truth. Dog can serve as a translational model and facilitate human cancer studies, like drug trials.

Previous work in our lab has shown that dog can serve as a model for the study of several kinds of human cancers at genomic level, since human and canine cancer share great molecular homologies.[2][3] For example, canine spontaneous head and neck squamous cell carcinomas and their human counterparts share similar genetic alterations, like sequence mutation patterns and alterations of pathways and genes.[2] Previous work has also shown that dog can be a model for the study of human mammary cancer. Both canine and human breast cancer have frequent CNAs. The genomic sites altered in human breast cancer also have high probability to be altered in canine breast cancer. The prevailed base substitution type in both species is C -T/G –A. Besides, the hypermutation is concurrent with DNA repair defection in both cancer types.[3] But there is little study to show the molecular homology between human and canine colorectal cancer. If dog can be shown to be a good model for human colorectal cancer, this will significantly benefit the study of human colon cancer.

We set out this study of the RNAseq data of canine colorectal cancer, in order to further compare genetic alterations of canine colorectal cancer to those of human. Also, to gain a better knowledge of dog colon cancer. This work mainly studied the expression characters of canine colorectal cancer by analyzing RNAseq data. We used 25 canine fresh frozen samples for RNA sequencing. The RNAseq data was mapped to dog canFam3 genome with Tophat. After mapping, we used cufflinks to resemble the transcripts and then calculated FPKM of each gene. The gene expression level of each sample described by FPKM were then analyzed with PCA and DEseq. Since colon is a special part with abundant microbes, the other unmapped reads were further mapped to bacteria genomes to detect potential pathogenic bacterium in these canine

colorectal cancer samples. The colon cancer samples were separated into 3 groups based on PCA analysis. By integrating the histology data and PCA analysis, the three groups are two different groups of carcinomas and one normal group. The DEseq analysis showed that, one of the carcinoma group has highly regulated cell cycle signaling and the other carcinoma group has highly regulated immune-response signaling. Future work can do more analysis based on these two groups to better characterize canine colorectal cancer. A lot of more work needs to be done to study the mutated genes, mutation rates and chromosomal translocations based on this work in order to explore potential molecular homology between canine and human colorectal cancer.

C1orf63 was found to be related to cell proliferation in our lab's previous work. Its expression was highly regulated in Taxol treated cells. Taxol is a plant alkaloid used in the treatment of several human carcinomas. Taxol binds to β-tubulin and stabilizes the microtubules. As a result, Taxol cause mitotic arrest and cell death. [24] At spindle assembly checkpoint (SAC), CDC20 and bubR1, Mad2 form mitotic checkpoint complex to inhibit APC, which is the anaphase promoting complex. APC can target cyclin B for degradation and thus the cell can enter anaphase. Once each kinetochore is attached to the mitotic spindle, the SAC proteins are displaced, which releases Cdc20. This allows APC to target cyclin B for degradation. But when Taxol is present and the microtubules are stabilized, the SAC cannot be displaced. So the cells were arrest in spindle assembly checkpoint. Cells have developed mechanisms to escape from this arrest. For example, premature degradation of cyclin B1 and direct inhibition of Cdk1 activity can help the cell to escape this arrest and progress to anaphase. This process is called mitotic slippage. This mitotic slippage can result in polyploidy and help cell to survival from mitotic cell death. [26] C1orf63 was found to respond to Taxol treatment in this study. C1orf63 may function in helping cells to escape from mitotic arrest caused by Taxol. From qPCR results,

we found that C1orf63 RNA transcription level was increased when cells were treated with Taxol. Immunofluorescence staining also showed that C1orf63 expression increased in Taxol treated cells.

MATERIALS AND METHODS

Tophat mapping

RNAseq reads were mapped to dog genome canFam3 with Tophat.[6] The basic procedure of Tophat is that the RNAseq reads were firstly aligned to canFam3 genome with Bowtie. Then the splice junctions (normally at/gt sites) between exons were found. After that, reads were mapped to possible splices. The script used is as follows: --library-type fr-firststrand -N 5 --read-edit-dist 5 -a 5 -i 50 -I 500000. The uniquely mapped reads is calculated by aligned pairs minus discordant pairs, then minus multiple pairs. These three numbers were got from tophat mapping report. Most of the samples have 60% uniquely mapped reads.

Cufflinks and Cuffdiff

Cufflinks[7] is used to assemble transcripts and estimate the expression level of each gene. The accepted_hits.bam file from Tophat mapping is used as input file for both cufflinks and cuffdiff. –u option is used to more accurately weight reads mapped to multiple positions. –G option is used to use only the transcripts in the annotation file and not assemble novel transcripts. Cufflinks assembles the transcripts based on Tophat mapping results, and estimates the relative abundance of each transcript based on how many reads mapped to it. The relative abundance is represented by FPKM[8], which is fragments per kilobase of transcript sequence per millions base pairs sequenced. The annotation file for cufflinks and cuffdiff is downloaded from UCSC. The raw count data from cuffdiff is used for following study. The script for cuffdiff is as follows: -N -T -b ../source/canFam3.fa --library-type fr-firststrand -u -c 5. The script for cufflinks is as follows: -G -b –u.

PCA and DEseq

Gene FPKM values were used for PCA[10] analysis and gene count values were used for DEseq[9] analysis. The genes have at least one value over 1 among the samples analyzed were used for PCA and DEseq analysis. For PCA, the FPKM values were normalized by log2 transformation. The heatmap for DEseq was generated by heatmap.2 R package [12]. Pairwise DEseq was carried out among the three sub groups. Significantly differently expressed genes among each sub groups were identified at FDR threshold 0.2. Then, we merge genes detected by each DEseq analysis among the three sub groups together. Then the raw count values for each gene were replace by FPKM value. And the DEseq heatmap was drawn using this data. The DEseq heatmap were unsupervised clustered on the gene level and on sample level. DAVID functional annotation tool was used to show enrichment of functional annotation of each cluster in heatmap against gene sets defined by GO_TERM annotation.

Inflammation study

The unmapped RNAseq reads were mapped to all the NCBI completed bacteria genomes (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). For the specie with multiple strains, only the strain with the longest genome was reserved.[5] The mapping was done by Bowtie2[13] with the following parameters: –local -D 20 -R 3 -N 1 -L 32 -i S,1,0.50. The mapping results contain unmapped reads, multiple mapped reads and uniquely mapped reads. Pathoscope[11] was used to assign multiple mapped reads to one genome. Then the uniquely mapped reads are used for following calculation. After mapping, the identified bacteria genomes are further filtered by method from "Zhang et al. Genome Biology (2015) 16:265 DOI 10.1186/s13059-015-0821-z".[5] Basically, to detect positive bacteria genome mapping, the genomic coverage is calculated. In this study, var1<=300 and var2<=0.0005[5] are set to determine positive bacteria

identification. The RPM for each genome is the total number of uniquely mapped reads to each genome per million total mapped reads of this sample.

Functional study of C1orf63

Taxol treatment of HCT116 cells

Taxol treatment experiments were carried out following a published method.[25] This experiment is done in HCT116 cell line. Taxol (Sigma) stock solution was prepared in DMSO to the concentration of 1mM. The stock solution was maintained at -20℃. Cells grow to $2.5x10^4$ cells/ml density were treated with various concentrations of Taxol at 37℃.

C1orf63 expression quantification by qRT-PCR

Cells with or without Taxol treatment were harvested at various incubation times. 5 million cells were used for RNA extraction using Qiagen RNeasy Plus Mini kit (Cat. No. 74134). Then, cDNA was synthesized from RNA template. The synthesized cDNA was used as template for qRT-PCR. The system was 2.5ul of 1:20 diluted cDNA, 500nM primer, 5ul 2X SYBR Green SuperMix(BioRad). Ct values were normalized to GAPDH by $RE_{Gene} = 100 \times 2^{-(Ct_{Gene} - Ct_{GAPDH})}$.

Immunofluorescence

Immunofluorescence study were performed with cell cultures. Primary antibody was against c1orf63. Images were taken with a Zeiss LSM 710 confocal microscope.

RESULTS

PCA analysis reveals three groups among all the sequenced samples

Gene expression profiles of canine colorectal cancers show great heterogeneity. So, PCA analysis was performed to classify the samples. Since FPKM value less than 1 will cause great difference in the PCA analysis but have little meaning in the analysis, so we chose those genes that have at least one FPKM value larger than 1 among all the samples to do the PCA analysis. After this filtering, there were 15652 genes left for the PCA analysis. From figure 1, the samples were separated into three groups. Although with so large amount of genes, the samples are not

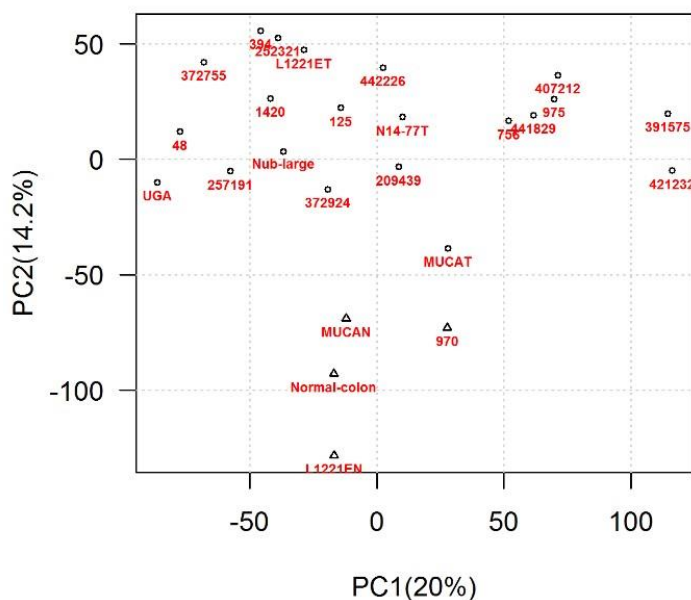

Fig.1 PCA analysis with all the genes that have at least 1 FPKM value larger than 1 over all samples.

grouped very well, with the primary component represents 20% of all the genes, and the secondary component represents 14.2% of all the samples, the three groups are still clear. Based on the first component, carcinoma group 1 is separated from carcinoma group 2. Based on the second component, tumors are separated from normal samples. Three normal samples, L1221EN, MUCAN and the normal colon sample, are grouped together on the PCA figure. Besides, the 970 tumor sample expressed like normal, so we classify it into normal group. L1221ET,394,372755,1420,252321,48,UGA,257191,442226,N14-77T,125,209439,Nub-large,372924,MUCAT is a group of carcinoma which is highly regulated in cell cycle related gene expression based on following study and is denoted as carcinoma group 1. 391575,407212,975,441829,756,421232 is a group of carcinoma which is highly regulated in immune response related gene expression based on following study and is denoted as carcinoma group 2.

Beside the total genes, we also do PCA analysis with several list of genes that have significant biological functions, including cell cycle genes, innate immune genes, metabolic genes, wound healing genes and chromatin genes. Figure 2 shows that the metabolic genes separate the samples into three groups largely resembling the first PCA analysis. This indicate that metabolic regulation is different between the two kinds of carcinomas. The results for other gene lists can be found in supplementary material.

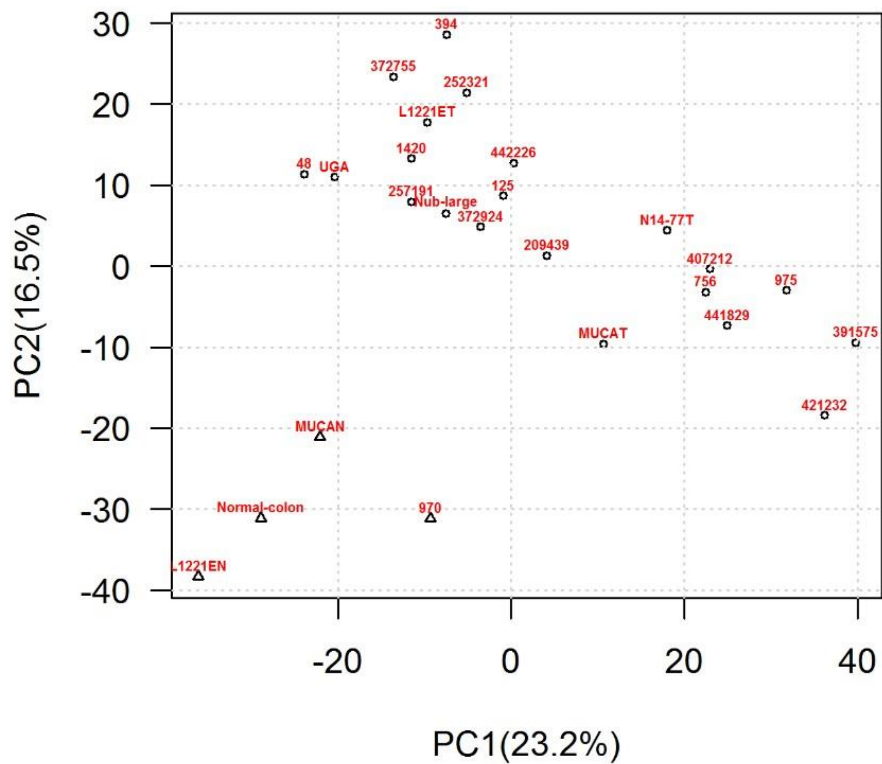

Fig2. PCA analysis of CRC samples with 1871 metabolic gene FPKM values. Total 1871 genes with at least 1 FPKM value larger than 1 over all samples are used for the analysis. Triangles represent normal samples and circles represent tumor samples. Metabolic genes separate the CRC samples into basically three groups. The normal sample group(including Normal, MUCAN, L1221EN, 970) and one carcinoma group(including 407212, 756, 441829, 391575, 975 and 421232) and another carcinoma group(including 394, 372755, 252321, L1221ET, 1420, 442226, 125, 257191, Nub-large, 372924, 48, 209439, MUCAT, N14-77T and UGA), which is basically in accordance with all gene PCA analysis.

DEseq analysis shows that one carcinoma group has highly regulated cell cycle signaling and the other carcinoma group has highly regulated immune-response signaling

Based on the PCA analysis, we want to further study the difference between the two carcinoma groups and also between the carcinoma groups and normal group. We did DEseq analysis to detect differently expressed genes among groups. After three pair-wise DEseq analysis, we merged all the detected genes together and draw an unsupervised heatmap. Then, the genes clustered together on the heatmap were uploaded on The Database for Annotation, Visualization and Integrated Discovery website (https://david.ncifcrf.gov/) to do functional annotation. The significantly enriched functions of each panel were indicated on the right of the figure. On the FDR0.2 level, inflammatory response is highly regulated in carcinoma group 2 and cell cycle regulation is highly regulated in carcinoma group 1. This propound the hypothesis that carcinoma group one is originated from cell cycle dysregulation while carcinoma group 2 is originated from inflammation.

Fig3. DEseq analysis heatmap. A heatmap shows gene expression profile in the three groups. Rows are genes and columns are samples. The expression level is represented by a color scale: dark blue for low expression levels and red for high expression levels. The genes are detected by DEseq with FDR less than 0.2. For each gene there is at least one FPKM >=1 among all the samples. The right panel shows the significantly enriched functions of each gene cluster indicated. The last column shows the P value.

Inflammation study shows that bacteria in the two carcinoma groups are different

Colon tissues are abundant with microbes. Chronical inflammation has been shown to be closely related to colon cancer.[23] Since DEseq analysis shows that the two carcinoma groups are different in immune response, we want to study if there are differences between the residential bacteria in the two carcinoma groups. Although the pathogen detection study shows that the carcinoma group1 and carcinoma group 2 have different bacteria, we are still uncertain whether carcinoma group2 are originated from chronical inflammation.

On the RNAsesq level, there are little bacteria detected, primarily because RNA is less stable than DNA and the trace amount of RNA of a special specie is hard to be detected. But on the other side, whatever is found in the RNAseq level should be considered more convenience to be present in the dog colon samples. From the RNAseq pathogen detection figure (fig4), we can see that Stenotrophomonas maltophilia, Comamonas testosterone, Pseudomonas are almost present in all colon samples. S. maltophilia is an uncommon bacteria and its infection have been associated to HIV infection, malignance, cystic fibrosis.[21] Comamonas testosteroni is commonly isolated from activated sludge. This organism is able to degrade a number of different compounds including chlorobenzenes. The deep relationship between its infection and cancer need to be further illustrated.

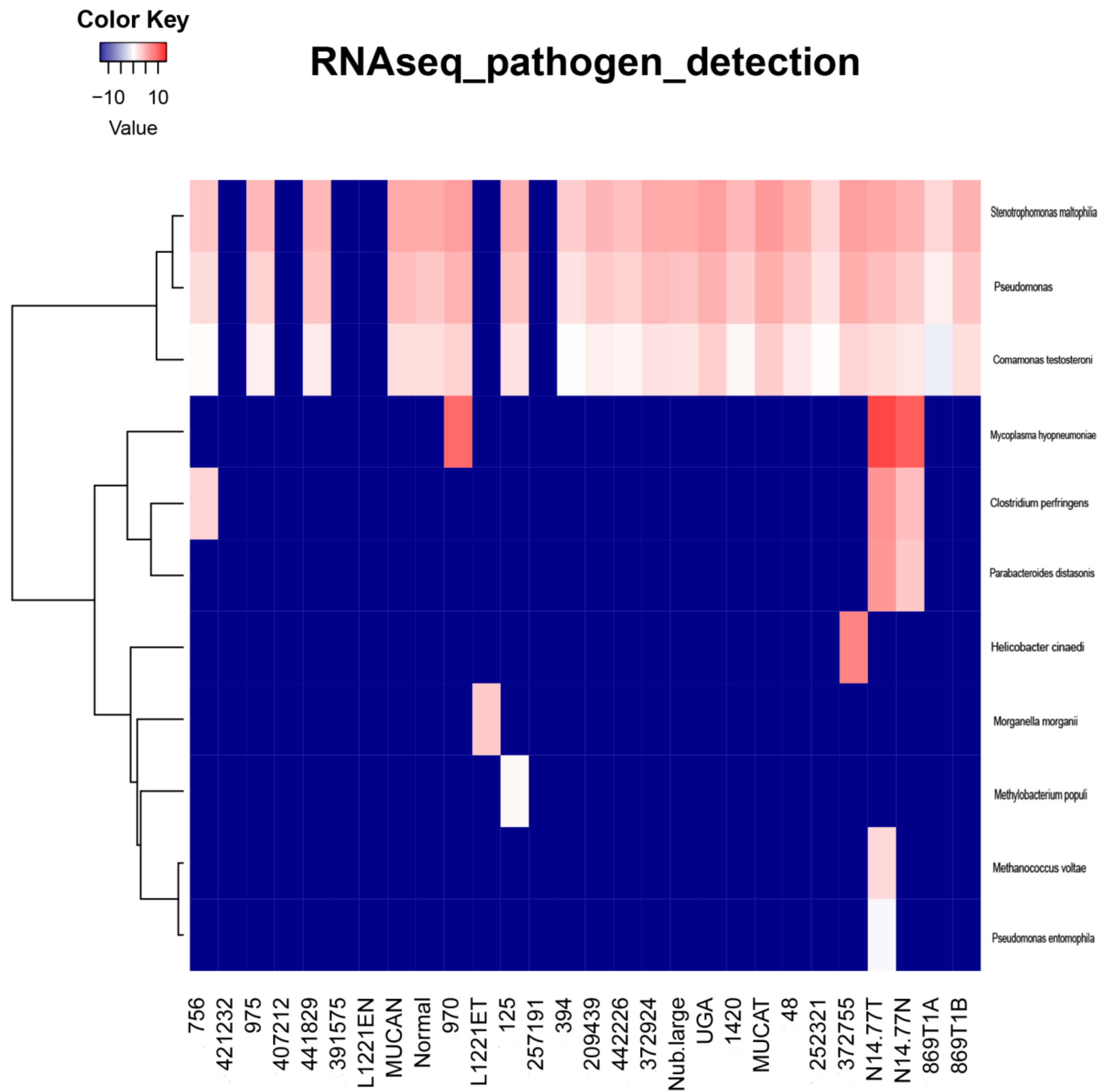

Fig4. Pathogen detection with

RNAseq data

C1orf63 expression is increased in Taxol treated cells

Previous work in our lab have found that C1orf63 may have functions in cell proliferation. To further explore the function of C1orf63, we did this experiment to study the role of C1orf63 in mitosis. Taxol is a plant alkaloid that can stabilize microtubules and arrest cells in spindle assembly checkpoint. We treated the cell cultures with Taxol at different concentrations for different time period. And then we use qPCR to analyze the expression level of C1orf63 under different conditions. From figure 5, we can see that the expression level of C1orf63 normalized by Gapdh is increased as the concentration and incubation time of Taxol increases. This response resembles GADD45A which is a growth arrest and DNA-damage-inducible protein encoding gene. This indicates that C1orf63 may have functions in response to cell stress and mitotic arrest.

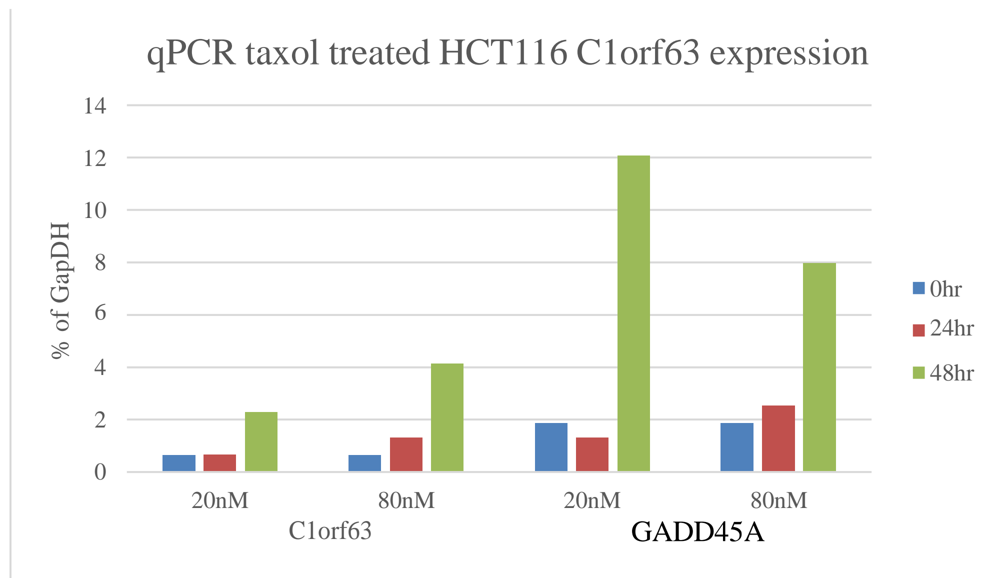

Fig5. C1orf63 expression level in Taxol treated HCT116. Y-axis is the expression level of C1orf63 and GADD45A represented by GAPDH normalized Ct value. 20nM and 80nM indicate the concentration of Taxol treatment. Blue bar is without Taxol treatment; red bar is Taxol treated for 24hr; green bar is Taxol treated for 48hr.

To show the localization of C1orf63 in Taxol treated cells, we did immunofluorescence staining with Taxol treated HCT116 cell cultures. From figure 6, we can see that since Taxol arrest the cells in mitosis metaphase, most of the cells have condensed chromosomes. The location of C1orf63, which is indicated by red dots, resembles centrosomes in a cell. After incubation with Taxol, HCT116 cells entered mitosis, which is shown like rounded up cells. It remained there for prolonged time without cell division, or cell death. Then the cells spread and appeared as flattened interphase cells with multiple nucleus. In these cells, C1orf63 were shown as multiple dots located alongside the chromosomes. This propose the hypothesis that C1orf63 may function in the centrosome and lead to polyploidy when the cells are under stress of Taxol. This will potentially help the cells to survive from mitotic cell death. But further studies are need to illustrate the detailed functions of C1orf63 in mitosis.

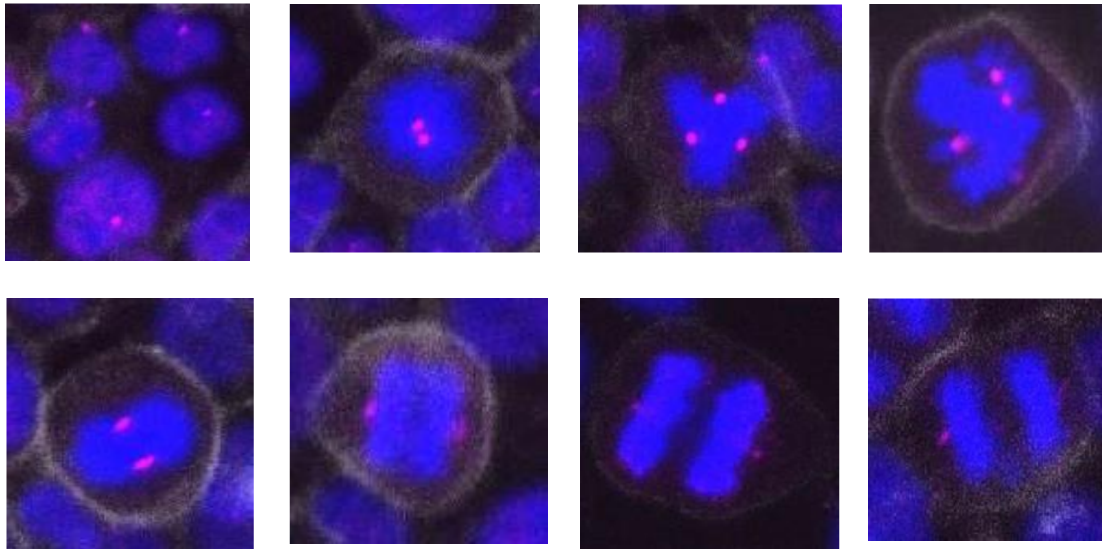

Fig6. Immunofluorescence staining of C1orf63. Red: C1orf63; blue: Dapi; white: E-cadherin. 40X.

DISCUSSION

This study analyzed the canine colorectal cancer RNAseq level data to show the gene expression characters of canine colorectal cancer. The PCA analysis shows that there are two kinds of carcinomas. From DEseq analysis, we can see one kind of the carcinomas has highly regulated immune response and the other has highly regulated cell cycle. The inflammation study shows that there are potential pathogens that can initiate cancer in these samples.

The most important limitation of the inflammation study is the volume of available sequenced bacteria genomes. We can easily imagine that there must be numerous pathogens that play fundamental roles in the causation of dog colon cancer but have not been sequenced. Inflammation study of canine colorectal cancer have its inherent significants in colon cancer research for the following reasons. 1. Colon have abundant microbes. The interaction between microbes will do benefits as well as damages to the colon tissue. Some normal bacterial infection may predispose colon to sever infections. Bacterial infection lead to cancer through two mechanisms: induce chronic inflammation and produce carcinogenic bacterial metabolites. [18] Mutagenic bacterial metabolites play important role in colon cancer. Bile salt have been shown to induce colon cell proliferation. Some compounds may be metabolized into mutagens by resident colonic bacteria. [18] 2. Pathogens have been known to be important causation of various kind of carcinomas. A well-known example of the connection between bacteria and cancer is the Helicobacter pylori infection. The inflammation of Helicobacter pylori infection will induce cell proliferation and produce mutagenic free radical which eventually cause gastric cancer.[19] 3. Besides, dog have more chances to be exposed to bacteria than human. So more

solid bacterial infection study of dog colon cancer is necessary to illustrate the connection between bacterial infection and the initiation of colon cancer. In addition, since bacterial infection can be prevented by antibiotics, the study of bacterial infection will have significant meaning for clinical therapeutics of human cancer. [18]

For future work, we can further study the human TCGA RNAseq data. Then compare this canine colorectal cancer expression profile to the human study to reveal the molecular homology between human and canine colorectal cancer. This will show if dog can be a potential model for the study of human colorectal cancer at the gene expression level. Besides, more molecular characters of canine colorectal cancer can be revealed by further study of this RNAseq data. For example, the somatic mutation rate, somatic gene alteration and microsatellite instability types. This will further illustrate the similarity between human and canine colorectal cancer, and corroborate dog as a model for human cancer study.

C1orf63 expression was increased in Taxol treated HCT116 cells. Immunofluorescence results also show that C1orf63 have similar cellular location as centrosome. This indicates that C1orf63 may have functions in mitosis. But the mechanism through which C1orf63 functions in mitotic arrest needs to be further explored.

Reference

1. *Nature* **487**, 330–337(19 July 2012) doi:10.1038/nature11252

2. Liu D, Xiong H, Ellis AE, Northrup NC,Dobbin KK, Shin DM, et al. (2015) Canine Spontaneous Head and Neck Squamous Cell Carcinomas Represent Their Human Counterparts at the Molecular Level. PLoS Genet 11(6): e1005277. doi:10.1371/journal.pgen.1005277.

3. Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez Jr CO, O'Regan RM, et al. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. Cancer Res. 2014;74:5045–56.

4. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005; 438(7069):803–19. Epub 2005/12/13.

5. Zhang et al. Genome Biology (2015) 16:265 DOI 10.1186/s13059-015-0821-z.

6. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* doi:10.1093/bioinformatics/btp1

7. Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg,John L Rinn& Lior Pachter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7, 562–578 (2012) doi:10.1038/nprot.2012.016

8. Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer & Barbara Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods - 5, 621 - 628 (2008) doi:10.1038/nmeth.1226

9. Simon Anders and Wolfgang Huber (2010): Differential expression analysis for sequence count data. Genome Biology 11:R106 doi: 10.1186/gb-2010-11-10-r106

10. Stacklies, W., Redestig, H., Scholz, M., Walther, D. and Selbig, J. pcaMethods -- a Bioconductor package providing PCA methods for incomplete data. Bioinformatics, 2007, 23, 1164-1167

11. Allyson L Byrd, Joseph F Perez-Rogers, Solaiappan Manimaran, Eduardo Castro-Nallar, Ian Toma, Tim McCaffrey, Marc Siegel, Gary Benson, Keith A CrandallEmail author and William Evan Johnson. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. BMC Bioinformatics201415:262 DOI: 10.1186/1471-2105-15-262

12. Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables. gplots: Various R Programming Tools for Plotting Data.

13. Ben LangmeadEmail author, Cole Trapnell, Mihai Pop and Steven L Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology200910:R25 DOI: 10.1186/gb-2009-10-3-r25

14. Liu W, Feng Z, Fang L, Zhou Z, Li Q, Li S, Luo R, Wang L, Chen H, Shao G, Xiao S. Complete genome sequence of Mycoplasma hyopneumoniae strain 168. J Bacteriol. 2011 Feb;193(4):1016-7. doi: 10.1128/JB.01305-10.

15. Siqueira FM, Thompson CE, Virginio VG, Gonchoroski T, Reolon L, Almeida LG, da Fonsêca MM, de Souza R, Prosdocimi F, Schrank IS, Ferreira HB, de Vasconcelos AT, Zaha A. New insights on the biology of swine respiratory tract mycoplasmas from a comparative genome analysis. BMC Genomics. 2013 Mar 14;14:175. doi: 10.1186/1471-2164-14-175.

16. Tajima, M., and T. Yagihashi. 1982. Interaction of Mycoplasma-Hyopneumoniae with the Porcine Respiratory Epithelium as Observed by Electron-Microscopy. Infection and Immunity 37:1162-1169.

17. Minion, F. C. 2002. Molecular pathogenesis of mycoplasma animal respiratory pathogens. Frontiers in Bioscience 7:d1410-1422.

18. Parsonnet J. Bacterial infection as a cause of cancer. Environ Health Perspect. 1995 Nov;103 Suppl 8:263-8.

19. Asaka M, Takeda H, Sugiyama T, Kato M. What role does Helicobacter pylori play in gastric cancer? Gastroenterology. 1997 Dec;113(6 Suppl):S56-60.

20. Li H, Limenitakis JP, Fuhrer T, Geuking MB, Lawson MA, Wyss M, Brugiroux S, Keller I, Macpherson JA, Rupp S, Stolp B, Stein JV, Stecher B, Sauer U, McCoy KD, Macpherson AJ. The outer mucus layer hosts a distinct intestinal microbial niche. Nat Commun. 2015 Sep 22;6:8292. doi: 10.1038/ncomms9292.

21. Chang, Ya Ting; Lin, Chun Yu; Chen, Yen Hsu; Hsueh, Po-Ren (2015-01-01). "Update on infections caused by Stenotrophomonas maltophilia with particular attention to resistance mechanisms and therapeutic options". Infectious Diseases: 893. doi:10.3389/fmicb.2015.00893

22. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe E, Holt RA. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome. 2013 May 15;1(1):16. doi: 10.1186/2049-2618-1-16.

23. Pesic M, Greten FR. Inflammation and cancer: tissue regeneration gone awry. Curr Opin Cell Biol. 2016 Aug 10;43:55-61. doi: 10.1016/j.ceb.2016.07.010.

24. Xiao H, Verdier-Pinard P, Fernandez-Fuentes N, Burd B, Angeletti R, Fiser A, Horwitz SB, Orr GA. Insights into the mechanism of microtubule stabilization by Taxol. Proc Natl Acad Sci U S A. 2006 Jul 5;103(27):10166-73. doi: 10.1073/pnas.0603704103.

25. Llovera L, Mansilla S, Portugal J. Apoptotic-like death occurs through a caspase-independent route in colon carcinoma cells undergoingmitotic catastrophe. Cancer Lett. 2012 Dec 29;326(1):114-21. doi: 10.1016/j.canlet.2012.08.001.

26. Burgess A, Rasouli M, Rogers S. Stressing mitosis to death. Front Oncol. 2014 Jun 4;4:140. doi: 10.3389/fonc.2014.00140.