Journal of Veterinary Science & Technology

Irizarry, et al., J Veterinar Sci Techno 2016, 7:3
DOI: 10.4172/2157-7579.1000322

Research Article Open Access

Leveraging Naked Mole Rat (*Heterocephalus glaber*) Comparative Genomics to Identify Canine Genes Modulating Susceptibility to Tumorigenesis and Cancer Phenotypes

Kristopher JL Irizarry[1,2*], **Natalie Punt**[1], **Randall Bryden**[1], **Joseph Bertone**[1] **and Yvonne Drechsler**[1]

[1]*College of Veterinary Medicine, Western University of Health Sciences, 309 East Second Street Pomona, California 91766-1854, USA*

[2]*The Applied Genomics Center Graduate College of Biomedical Sciences, Western University of Health Sciences, 309 East Second Street Pomona, California 91766-1854, USA*

[*]**Corresponding author:** Kristopher JL Irizarry, College of Veterinary Medicine, Western University of Health Sciences, 309 East Second Street Pomona, California 91766-1854, USA, Tel: 909-584-2570; E-mail: kirizarry@westernu.edu

Rec date: Dec 09, 2015; **Acc date:** Mar 29, 2016; **Pub date:** Mar 31, 2016

Abstract

We utilized a comparative genomics approach to analyze a core set of tumorigenesis orthologs among human, mouse, dog and naked mole rat. The analysis identified cancer orthologs that are both conserved and divergent between dog and the cancer resistant species naked mole rat. These tumorigenesis orthologous are associated with phenotypes that modulate cancer susceptibility, cardiac development, craniofacial development, brain development, skeletal development, and immune function, to name a few. This bioinformatics approach employed a variety of literature mining tools to further uncover relationships between the tumorigenesis orthologs. Together, these results shed light on the relationship between breed formations, breed associated morphological traits and breed associated susceptibility to tumorigenesis. These findings support the use of a comparative genomic analysis between species with dramatically different disease phenotypes as a gene discovery tool. A total of 146 proteins coding SNPs were identified in these tumorigenesis orthologs representing missense variations, frame shift variations and nonsense variations. The genes identified in this study can serve as a list of candidates for subsequent laboratory and clinical study. Furthermore, the identification of SNPs impacting the primary structure of the tumorigenesis orthologs may provide clues about the basis of cancer susceptibility between dog breeds.

Keywords: *Heterocephalus glaber*; Tumorigenesis; Urothelial carcinoma; Canine cancer; Mammalian; Dog

Introduction

Cancer is a complex disease that occurs in humans and animals. It is estimated that approximately 1.5 million humans and 4 million dogs are diagnosed with cancer each year [1]. Many cancer types occur in both humans and dogs including bladder, head, lung, mammary, neck, and prostate carcinomas; leukemia; non-Hodgkin lymphoma; melanoma; soft tissue sarcomas; and osteosarcoma [2]. Records from the Swiss Canine Cancer Registry dating from 1955 to 2008, in which tumors were classified using the International Classification of Oncology for Humans (ICD-O-3) by anatomical location, tumor type and malignancy, were recently analyzed to assess the distribution of cancer in dogs [3]. The most common tumors identified in these records (n=67,943) included adenoma (32.6%), neoplasia of stroma (9.6%), fibrosarcoma (8.5%), mast cell sarcoma (6%), blood vessel neoplasia (5.5%), lipoma (5.5%), soft tissue sarcoma (4.4%), epithelial tumor (4.4%), lymphoma (4.4%), unclassified neoplasm (4%), gonadal tumor (1.6%), skeletal tumor (1.5%), and histiocytic neoplasm (1.1%). The anatomical location of these tumors included skin (32.3%), mammary gland (20.5%), unclassified location (12.8%), soft tissues (11.9%), gastrointestinal tract (7.5%), male sexual organs (3.9%), respiratory tract and intrathoracic organs (2.1%), blood and hematopoietic system (2.1%), bones/joints/articular cartilage (1.6%), endocrine glands (1.3%), oral cavity and pharynx (1.2%), other female sexual organs (0.9%), urinary organs (0.5%), central nervous system (0.4%), lymph nodes (0.4%), retroperitoneum and peritoneum (0.2%), eyes and perception organs (0.2%), as well as the peripheral nervous system (0.1%).

Comparative oncology studies focusing on specific cancers have identified shared aspects of histology and pathophysiology between dogs and humans. For example, canine lymphoma and human non-Hodgkin's lymphoma are both heterogeneous lymphoid diseases comprised of a variety of different cells exhibiting diverse biological behavior. Moreover, breed specific prevalence for B-cell versus T-cell lymphoma has been observed, with Siberian Husky exhibiting 100% T-cell lymphoma, Boxer exhibiting 65% T-cell lymphoma, and Golden Retriever having almost equal prevalence of B-cell and T-cell lymphoma while Cocker Spaniel and Doberman Pinscher have greater than 90% prevalence of B-cell lymphoma [4].

Another study investigated the similarity in gene expression patterns between human urothelial carcinoma samples and canine urothelial carcinoma samples. The results indicated that 436 genes exhibited altered expression in urothelial carcinoma compared to normal urothelial tissue in both dogs and humans [5]. Similarly, characterization of DNA copy number variations associated with osteosarcoma identified similar chromosomal abnormalities within the same genes in both human and canine cancer samples [6]. Moreover, genetic variation, within particular locations of certain genes has been associated with susceptibility to similar tumors in both humans and dogs. For example, single nucleotide polymorphisms in BRCA1 and BRCA2 have been associated with a four-fold relative risk of mammary tumors in dogs [7] paralleling the identification of the genetic variation within the same genes in human breast cancer [8].

As cancer arises through the dysregulation of the cell cycle, many genes implicated in cancer are functionally conserved across mammals. One example is the tumor suppressor p53, for which inactivating mutations are the most frequently observed molecular defects in human cancer [9]. The p53 tumor-suppressor pathway regulates cell cycle progression, DNA repair, and apoptosis in vertebrate cells and subsequently plays a pivotal role in tumorigenesis in the face of DNA damage. A single nucleotide polymorphism within codon 72 of human p53 encodes either a proline or an arginine at position 72 within the protein, resulting in allele-associated variation in the ability of the protein to promote apoptosis [10]. An earlier study in dogs identified numerous mutations in p53 within malignant lymphoma, monocytic leukemia, rhabdomyosarcoma, colon cancer, and osteosarcoma [11]. Therefore, it is not surprising that studies of mouse models of cancer have exploited various mutations of p53 to further elucidate the role of this gene in tumorigenesis [12].

Much as comparative anatomy and comparative physiology provide a context for understanding the mechanisms underlying inter-species variation in form and function, comparative genomics offers a similar approach for understanding conserved and divergent genetic mechanisms underlying phenotypes of interest. An example of a comparative genomics approach to identify cancer genes is the comparison of the Tasmanian Devil genome with the human genome in an attempt to identify cancer- related genes that might underlie the unique devil facial tumor disease of transmissible cancer [13]. The results of this study identified a number of orthologs of human cancer genes that exhibited genetic sequence variation within the Devils associated with devil facial tumor disease compared to those without disease.

A more recent example of a comparative genomics approach to identify anti-cancer mechanisms leveraged the genome of the African elephant to investigate the genomic basis of reduced tumorigenesis within this species [14]. The results of this comparative study identified approximately 20 duplications of the p53 tumor-suppressor gene in the African elephant genome, which likely contribute to a decreased rate of tumorigenesis in the elephant.

The use of model organisms to elucidate genetic mechanisms has included the fruit fly, the worm, and the mouse, to name a few. The mouse is the most widely studied mammalian laboratory model. Gene knockout technology has resulted in the production of mouse strains with inactivating mutations in over one third of the genes encoded in the mouse genome [15]. The International Mouse Phenotyping Consortium (IMPC), (a collaborative functional genomics effort between laboratories in America, Germany, United Kingdom, France, Canada, China and Japan) has characterized phenotype data for 2000 mouse genes and plans to have 5000 genes characterized by Ring et al. [16]. The Mouse Genome Database (MGD) provides a central repository for mouse functional genomics data and resources including phenotype annotations for mouse genes using the Gene Ontology and Mammalian Phenotype Ontology [17]. The MGD contains 24,613 mouse genes mapped to 17,055 human genes, of these, 12,619 genes have one or more phenotypic alleles and, in total, 52,570 alleles (multiple alleles in the same gene) are associated with phenotypic annotation. Within this set of genomic data, 9225 mouse genes are associated with targeted alleles, such as gene knock outs. Additionally, a total of 24,605 protein coding mouse genes have at least one functional gene ontology (GO) annotation resulting in a total of 291,605 gene-GO annotations across all mouse genes.

Unlike the mouse, and most mammals, the naked mole rat has not been observed with spontaneous cancer, additionally, naked mole rat cells exhibit resistance to tumors when transduced with oncogenic genes that promote cancer in other mammalian species [18]. In contrast to mice and rats, which have relatively short-life spans, the naked mole rat has a documented lifespan over 30 years, and represents a mammal with a unique anti-cancer phenotype [19].

Here we report a comparative genomics approach to identify genes in the dog that are likely to be associated with susceptibility and resistance to tumorigenesis. Our hypothesis is that genes associated with mouse tumorigenesis phenotypes that are least similar between dog and naked mole rat are possibly enriched for genes that underlie variation in the tumorigenesis phenotypes between these two species. Specifically, genes exhibiting the most divergence between dog and naked mole rat may represent genes that are modifier genes (i.e., enhancers or suppressors of tumorigenesis) which may decrease the incidence, age-of onset, and/or progression of tumorigenesis within and/or between dog breeds.

In order to identify genes in the dog that may be associated with susceptibility to tumorigenesis, we identified a set of orthologous genes across human, mouse, dog and naked mole rat for which mouse alleles are associated with tumorigenesis. We then identify the most and least identical protein sequences between the dog and the naked mole rat and characterize the functional genomic annotation associated with these subsets of tumorigenesis genes.

Materials and Methods

Protein sequences

Canis familiaris (dog), *Homo sapiens* (human), and *Mus musculus* (mouse) protein coding sequences were downloaded from the Ensembl (http://www.ensembl.org) vertebrate genomics repository [20]. Protein coding sequences for Heterocephalus glaber (naked mole rat) were obtained from the protein sequence data (http://www.ncbi.nlm.nih.gov) available at NCBI [21].

Ortholog detection

Orthologs were detected (Figure 1A) using the BLAST reciprocal best hit method [22]. All six possible BLAST pairwise species sequence comparisons were completed in both directions (speciesX vs speciesY and speciesY vs. SpeciesX) and the resulting tab-delimited output files were loaded into the open source MySQL relational database (https://www.mysql.com/). Highest scoring BLAST hit for each query sequence, in each directional pair-wise species comparison, was identified and loaded into database tables (e.g.,: bestHits_speciesX_vs_speciesY, bestHits_speciesY_vs_speciesX, bestHits_speciesY_vs_speciesZ, bestHits_specieZ_vs_speciesY). The resulting 12 database tables (speciesX_vs_speciesY and speciesY_vs_speciesX for each of six pair-wise comparisons) were used to identify reciprocal best hits. Specifically, proteinA in speciesX was considered an ortholog of proteinA' in speciesY if and only if the best BLAST hit for query proteinA in species X is proteinA' in speciesY AND the best BLAST hit for query proteinA' in species Y is proteinA in speciesX. Reciprocal best hits for each pair of species were identified and loaded into an ortholog_speciesX_and_speciesY database table using an SQL query of the form:

CREATE TABLE orthologs_speciesX_and_speciesY (

```
forward_queryId VARCHAR(32) NOT NULL,
forward_subjectId VARCHAR(32) NOT NULL,
forward_eValue FLOAT NOT NULL,
forward_bitScore FLOAT NOT NULL,
reverse_queryId VARCHAR(32) NOT NULL,
reverse_subjectId VARCHAR(32) NOT NULL,
reverse_eValue FLOAT NOT NULL,
reverse_bitScore FLOAT NOT NULL,
INDEX (forward_queryId),
INDEX (forward _subjectId),
INDEX (forward _eValue),
INDEX (forward _bitScore),
INDEX (reverse_queryId),
INDEX (reverse _subjectId),
INDEX (reverse _eValue),
INDEX (reverse _bitScore)
);
INSERT INTO orthologs_speciesX_and_speciesY select
```

t1.queryId, t1.subjectId, t1.eValue, t1.bitScore, t2.queryId, t2.subjectId, t2.eValue, t2.bitScore from bestBlastHits_speciesX_vs_speciesY_forward t1, bestBlastHits_speciesY_vs_speciesX_reverse t2 where t1.queryId=t2.subjectId and t1.subjectId=t2.queryId;

Identification of shared orthologs

Orthologous protein sequences shared across the four species (Figure 1A) was identified with an SQL query that joined the six pair-wise ortholog database tables. The resulting set of shared orthologs was loaded into a separate database table for subsequent ease of access. This data set served as the foundation for downstream analyses of the data.

Identification of orthologous tumorigenesis proteins

Mouse genomic annotation maintained in the Mammalian Phenotype Browser [17] (Eppig et al.) was obtained from Mouse Genome Informatics (http://www.informatics.jax.org/) repository. Mouse genes annotated with the 'tumorigenesis' phenotype were identified and used to select the corresponding orthologs from among the total set of shared orthologs across the four species (Figure 1A).

High-throughput literature mining

Literature mining (Figure 1B) was accomplished the PubMatrix (http://pubmatrix.grc.nia.nih.gov/) automated interface for querying PubMed [23,24]. PubAtlas (http://www.pubatlas.org/) was used to determine co-occurrence frequencies for pairs of query terms within pubmed abstracts [25]. The PubTator bioinformatics resource (http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator) was used to mine specific categories of biomedical knowledge from more than 2500 specific pubmed abstracts [26]. The tab-delimited output from PubTator was further mined to extract the specific terms identified in the pubmed abstracts corresponding to diseases, phenotypes, disorders, genes, proteins, and small molecules. The mined PubTator output was used to generate word cloud visualizations (https://www.jasondavies.com/wordcloud/) associated with specific genes and particular biological search terms. The word cloud visualization tool was configured to scale on the order of log (number of words) and to consider each line a single word/term.

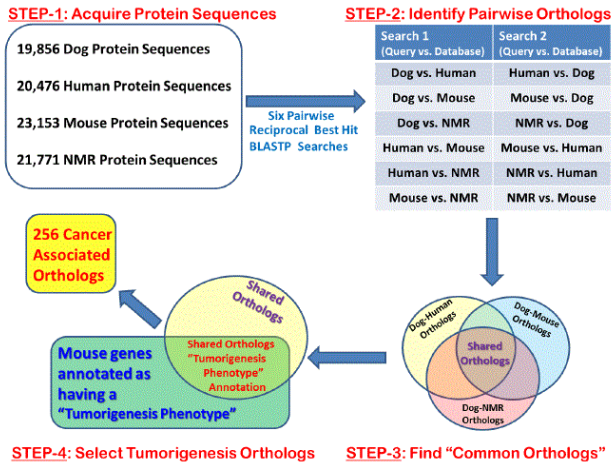

Figure 1A: Graphical Representation of Bioinformatics Method for Identifying Tumorigenesis Orthologs. Graphical representation of the four steps involved in identifying the tumorigenesis orthologs between human, mouse, dog and naked mole rat. First protein coding sequences from each of the four species was obtained. Second, Pairwise bi-directional BLAST comparisons were performed for each of the 6 pairs of species (note that bidirectional mapping resulted in 12 total pairwise BLAST analyses). Third, the set of shared one-to-one ortholgs was identified between the four species (using the three pair-wise ortholog mappings: dog-human, dog-mouse, and dog-naked mole rat. Fourth, the set of shared orthologs overlapping with a core set of mouse genes annotated experimentally as tumorigenesis were identified and labelled the 'tumorigenesis orthologs'.

Functional enrichment

The set of tumorigenesis orthologs shared across the species was analyzed to identify statistically significant annotation enrichments within the most conserved (top 30%) and least conserved (bottom 30%) proteins (Figure 1B). Gene ontology enrichment analysis was carried out with the GOrilla (http://cbl-gorilla.cs.technion.ac.il/) analysis and visualization tool [27]. Phenotype enrichment was assessed using MamPhea [28]. MamPhEA: a web tool for mammalian phenotype enrichment analysis. Analysis resource (http://evol.nhri.org.tw/phenome/index.jsp?platform=mmus) to identify mammalian phenotypes enriched within the top 30% conserved and bottom 30% conserved proteins compared to the rest of the genome. The phenotype analysis was performed using the Mus musculus phenotypes and fisher's exact test across all categories of mammalian phenotypes, not just the tumorigenesis phenotypes. Reported p-values were adjusted for multiple comparisons, using the Banjamini correction for multiple comparisons, to limit false positives. Subsequently adjusted p-values exhibit larger values than the non-adjusted p-values. The Database for Annotation, Visualization and

Integrated Discovery [29]. (DAVID) (https://david.ncifcrf.gov/) was used to identify statistically enriched Biocarta Pathways in the tumorigenesis orthologs.

SNP analysis

SNPs in shared orthologous tumorigenesis proteins were obtained from the dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) database [30] using boolean queries designed to specifically identify non-synonymous SNPs within the protein coding regions (Figure 1B). Correlation analysis was performed to assess the relationship between protein sequence identity (averaged across all four species for each shared tumorigenesis ortholog) and the number of SNPs reported in dbSNP for Canis familiaris (dog), Homo sapiens (human), and Mus musculus (mouse). Graphs and correlation coefficients were generated with Microsoft Excel.

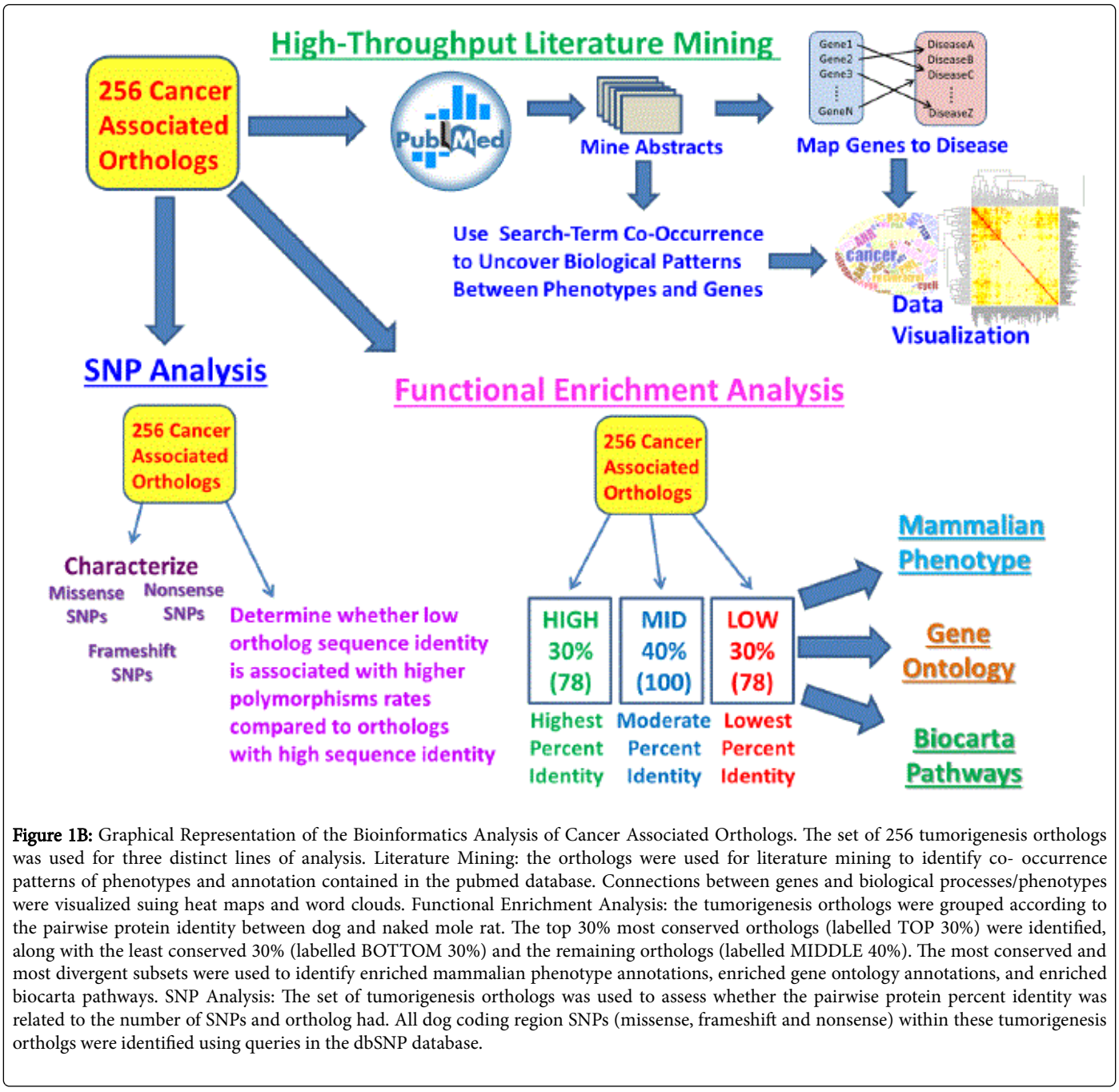

Figure 1B: Graphical Representation of the Bioinformatics Analysis of Cancer Associated Orthologs. The set of 256 tumorigenesis orthologs was used for three distinct lines of analysis. Literature Mining: the orthologs were used for literature mining to identify co- occurrence patterns of phenotypes and annotation contained in the pubmed database. Connections between genes and biological processes/phenotypes were visualized suing heat maps and word clouds. Functional Enrichment Analysis: the tumorigenesis orthologs were grouped according to the pairwise protein identity between dog and naked mole rat. The top 30% most conserved orthologs (labelled TOP 30%) were identified, along with the least conserved 30% (labelled BOTTOM 30%) and the remaining orthologs (labelled MIDDLE 40%). The most conserved and most divergent subsets were used to identify enriched mammalian phenotype annotations, enriched gene ontology annotations, and enriched biocarta pathways. SNP Analysis: The set of tumorigenesis orthologs was used to assess whether the pairwise protein percent identity was related to the number of SNPs and ortholog had. All dog coding region SNPs (missense, frameshift and nonsense) within these tumorigenesis ortholgs were identified using queries in the dbSNP database.

Results

Identification human, mouse, dog and naked mole rat tumorigenesis orthologs

Protein coding sequences from human (20,476), mouse (23,153), dog (19,856) and naked mole rat (21,771) were used to identify the set of shared orthologous protein sequences between all four species (Figure 1A). A total of 15,137 orthologous proteins were identified between human and naked mole rat. Similarly, 15,044 orthologs were identified between mouse and naked mole rat while just 14,405 orthologs were detected between dog and naked mole rat. The greatest number of orthologs were identified between human and mouse followed by human and dog (17,514) and then by mouse and dog (16,997). Mouse protein coding genes associated with phenotype annotation were filtered and those annotated as 'Tumorigenesis Phenotype' were identified. Subsequently, the set of 256 mouse tumorigenesis protein coding genes corresponding to one-to-one orthologs across human, mouse, dog and naked mole rat were ultimately selected as the tumorigenesis orthologs.

Characterization of protein sequence identity across the tumorigenesis orthologs

Amino acid sequence identity was calculated for each tumorigenesis ortholog between each specific pair of species (i.e., human vs. mouse, human vs. dog, human vs. naked mole rat, mouse vs. naked mole rat, dog vs. naked mole rat, mouse vs. dog). The resulting set of 1536 pairwise- protein identity scores are shown in Table 1. To better visualize the sequence identity relationships within these protein coding tumorigenesis genes, the data ordered by descending protein identity between dog and naked mole rat and used to generate a heat map (Figure 2). Each column in the heat map corresponds to a specific pair of species and each row of the heat map corresponds to a specific tumorigenesis ortholog. Pair-wise amino acid identity is represented by colors, with dark red corresponding to the highest sequence identity, white corresponding to moderate identity and blue representing the lowest amino acid identity between pairs of orthologs.

Dog Gene ID	Symbol	hum2nmr	mou2nmr	dog2nmr	hum2mou	mou2dog	hum2dog	
ENSCAFG00000001094	SMAD5	99.78	98.92	100.00	98.71	98.92	99.78	78 genes TOP 30%
ENSCAFG00000014105	SMARCB1	100.00	100.00	100.00	100.00	100.00	100.00	78 genes TOP 30%
ENSCAFG00000030297	GRB2	100.00	99.00	100.00	99.30	99.00	100.00	78 genes TOP 30%
ENSCAFG00000005204	CTNNB1	99.61	99.49	99.49	99.87	99.74	99.87	78 genes TOP 30%
ENSCAFG00000013493	PCYT1B	98.64	99.19	99.46	98.10	99.19	98.37	78 genes TOP 30%
ENSCAFG00000013499	RNF2	99.40	99.70	99.41	99.70	99.11	98.82	78 genes TOP 30%
ENSCAFG00000000164	SMAD4	99.09	98.73	99.31	98.37	98.63	98.63	78 genes TOP 30%
ENSCAFG00000013332	ATF2	99.00	98.88	99.16	99.18	99.59	99.38	78 genes TOP 30%
ENSCAFG00000028848	JUNB	99.16	90.91	99.16	92.80	91.64	94.52	78 genes TOP 30%
ENSCAFG00000007659	SMAD1	99.35	98.92	99.14	99.14	98.49	99.35	78 genes TOP 30%
ENSCAFG00000000068	BCL2	92.00	90.79	99.02	86.61	98.04	99.05	78 genes TOP 30%
ENSCAFG00000001378	MAPK14	98.89	98.33	98.61	99.44	99.17	99.72	78 genes TOP 30%
ENSCAFG00000003996	FGF2	98.51	99.25	98.51	94.84	94.84	95.06	78 genes TOP 30%
ENSCAFG00000013269	ID3	97.48	96.00	98.32	97.00	96.00	99.16	78 genes TOP 30%
ENSCAFG00000007869	ATP5A1	98.00	97.93	98.31	97.74	98.12	97.74	78 genes TOP 30%
ENSCAFG00000003783	FOXO3	97.43	95.29	98.29	93.77	96.20	98.67	78 genes TOP 30%
ENSCAFG00000002744	SPTBN1	98.56	98.05	98.28	97.00	99.00	98.77	78 genes TOP 30%
ENSCAFG00000000216	PGGT1B	98.18	96.97	97.88	96.82	96.02	97.88	78 genes TOP 30%
ENSCAFG00000012307	NF2	97.76	96.21	97.59	98.32	97.99	99.49	78 genes TOP 30%
ENSCAFG00000017388	SMAD3	97.59	97.59	97.34	100.00	99.76	99.76	78 genes TOP 30%
ENSCAFG00000004487	ERCC2	97.23	97.10	97.23	97.50	97.23	97.76	78 genes TOP 30%
ENSCAFG00000016077	BMPR1A	97.18	96.24	97.18	97.93	97.74	98.87	78 genes TOP 30%
ENSCAFG00000017040	FOS	96.05	93.16	97.11	93.68	93.95	97.11	78 genes TOP 30%

ENSCAFG00000005008	MGAT5	97.44	94.20	97.03	97.17	96.36	98.79	78 genes TOP 30%
ENSCAFG00000029853	CDKN1A	78.18	95.38	97.03	77.00	74.00	78.44	78 genes TOP 30%
ENSCAFG00000015695	PRKCH	97.62	97.17	97.02	97.80	97.21	97.66	78 genes TOP 30%
ENSCAFG00000000157	DCC	97.55	96.64	97.01	96.48	96.26	97.46	78 genes TOP 30%
ENSCAFG00000004567	PRDX1	96.98	93.00	97.00	95.00	95.00	98.99	78 genes TOP 30%
ENSCAFG00000001310	PPARD	66.00	68.10	96.98	93.22	90.80	95.00	78 genes TOP 30%
ENSCAFG00000013371	ETV6	96.48	84.00	96.92	88.79	88.18	97.36	78 genes TOP 30%
ENSCAFG00000001125	EP300	95.58	95.50	96.83	93.93	93.94	94.35	78 genes TOP 30%
ENSCAFG00000015545	SUV39H1	96.80	94.09	96.80	95.39	95.39	99.27	78 genes TOP 30%
ENSCAFG00000010651	TUSC2	90.91	87.00	96.72	93.00	100.00	98.36	78 genes TOP 30%
ENSCAFG00000010807	PDCD4	96.48	96.48	96.72	96.80	97.66	98.00	78 genes TOP 30%
ENSCAFG00000010740	GNAI2	97.18	96.90	96.62	98.31	97.46	98.59	78 genes TOP 30%
ENSCAFG00000010034	RYR2	97.07	96.09	96.47	96.88	96.21	98.11	78 genes TOP 30%
ENSCAFG00000014185	MEN1	87.00	95.78	96.42	96.22	96.58	98.36	78 genes TOP 30%
ENSCAFG00000007614	CCNE1	83.00	81.70	96.21	80.46	83.20	87.15	78 genes TOP 30%
ENSCAFG00000009596	RUNX1	96.00	94.53	96.17	95.81	95.16	98.24	78 genes TOP 30%
ENSCAFG00000012290	SFN	97.01	96.58	96.15	97.86	97.86	98.29	78 genes TOP 30%
ENSCAFG00000006055	ATP2C1	96.73	94.78	96.00	96.39	95.28	98.00	78 genes TOP 30%
ENSCAFG00000012332	H2AFX	96.00	95.00	96.00	96.69	97.52	99.00	78 genes TOP 30%
ENSCAFG00000012639	IQGAP1	96.58	95.67	95.98	96.00	95.00	96.98	78 genes TOP 30%
ENSCAFG00000018424	TOM1L2	97.24	94.87	95.87	94.87	92.72	96.85	78 genes TOP 30%
ENSCAFG00000001922	MAD2L1	97.09	93.00	95.63	94.00	94.00	97.56	78 genes TOP 30%
ENSCAFG00000017192	HMMR	80.95	94.30	95.27	95.42	95.15	96.73	78 genes TOP 30%
ENSCAFG00000019257	HIC1	80.95	94.30	95.27	95.42	95.15	96.73	78 genes TOP 30%
ENSCAFG00000001672	LEP	70.83	91.78	95.23	93.10	91.87	96.11	78 genes TOP 30%
ENSCAFG00000017855	SPARC	95.38	89.77	95.05	90.43	90.76	97.36	78 genes TOP 30%
ENSCAFG00000009654	RBL2	95.72	90.05	95.01	91.70	91.10	97.01	78 genes TOP 30%
ENSCAFG00000003821	ITGB1	92.98	93.98	94.99	92.48	94.49	95.36	78 genes TOP 30%
ENSCAFG00000019251	CREBBP	95.71	88.81	94.83	96.54	94.81	94.15	78 genes TOP 30%
ENSCAFG00000005260	SPRY2	96.51	93.99	94.60	96.20	94.30	97.14	78 genes TOP 30%
ENSCAFG00000010551	CHEK1	96.01	94.12	94.53	93.07	91.16	93.68	78 genes TOP 30%
ENSCAFG00000005166	HINT1	96.63	80.00	94.38	80.00	94.00	96.83	78 genes TOP 30%
ENSCAFG00000015670	PTEN	99.26	99.01	94.28	99.75	94.78	95.02	78 genes TOP 30%
ENSCAFG00000009156	IRF4	95.34	91.35	94.05	92.24	91.22	93.89	78 genes TOP 30%
ENSCAFG00000005965	FOXO1	94.02	95.45	94.02	90.58	88.89	93.82	78 genes TOP 30%
ENSCAFG00000023416	APEX1	94.34	56.00	94.00	73.60	64.40	78.00	78 genes TOP 30%

ENSCAFG00000006948	ALOX15	94.34	56.00	94.00	73.60	64.40	78.00	78 genes TOP 30%
ENSCAFG00000006742	CDX2	97.99	89.74	93.95	91.53	87.58	96.18	78 genes TOP 30%
ENSCAFG00000016007	KRT10	92.90	91.94	93.87	91.94	93.87	94.19	78 genes TOP 30%
ENSCAFG00000005449	TGFBR2	88.78	87.76	93.78	91.39	90.16	93.51	78 genes TOP 30%
ENSCAFG00000002454	XPA	94.23	88.00	93.75	86.00	85.00	95.19	78 genes TOP 30%
ENSCAFG00000018207	CABLES1	94.46	92.15	93.72	81.01	79.46	91.60	78 genes TOP 30%
ENSCAFG00000006598	MCM4	97.74	95.86	93.61	95.13	91.83	93.96	78 genes TOP 30%
ENSCAFG00000003247	MTF1	95.49	91.93	93.16	92.44	89.94	94.42	78 genes TOP 30%
ENSCAFG00000008853	TIAM1	95.29	93.78	93.15	95.47	94.41	95.73	78 genes TOP 30%
ENSCAFG00000004216	PTF1A	55.00	55.00	93.14	86.02	77.36	79.00	78 genes TOP 30%
ENSCAFG00000009421	MMP2	94.11	95.47	93.05	95.62	92.73	94.29	78 genes TOP 30%
ENSCAFG00000014345	FN1	92.38	92.40	92.82	91.44	92.21	94.50	78 genes TOP 30%
ENSCAFG00000015718	HIF1A	90.83	88.78	92.81	89.02	90.87	95.32	78 genes TOP 30%
ENSCAFG00000007150	ERCC8	92.93	87.91	92.68	89.67	90.18	94.44	78 genes TOP 30%
ENSCAFG00000007863	ATR	100.00	89.90	92.44	90.37	90.89	94.79	78 genes TOP 30%
ENSCAFG00000019869	RALGDS	87.99	83.80	92.44	89.70	92.42	93.17	78 genes TOP 30%
ENSCAFG00000008696	RBL1	92.04	89.13	92.42	90.45	91.01	95.04	78 genes TOP 30%
ENSCAFG00000016656	AR	90.60	89.11	92.26	83.51	86.69	86.99	78 genes TOP 30%
ENSCAFG00000012298	FES	92.94	91.12	92.21	90.63	90.02	94.04	78 genes TOP 30%
ENSCAFG00000015763	BECN1	90.97	92.19	92.19	98.00	97.99	97.56	100 GENES MIDDLE 40%
ENSCAFG00000016011	EPHA2	92.13	90.76	92.04	93.19	93.04	95.21	100 GENES MIDDLE 40%
ENSCAFG00000003373	DGKI	62.00	94.14	92.00	95.99	96.97	95.00	100 GENES MIDDLE 40%
ENSCAFG00000020031	WWOX	95.80	95.10	92.00	94.90	93.18	95.00	100 GENES MIDDLE 40%
ENSCAFG00000006088	ING1	58.87	91.00	91.88	88.00	93.00	65.81	100 GENES MIDDLE 40%
ENSCAFG00000014749	ANXA7	89.47	90.45	91.85	92.45	92.42	92.42	100 GENES MIDDLE 40%
ENSCAFG00000023631	ZBTB33	91.70	86.83	91.85	86.35	86.80	93.30	100 GENES MIDDLE 40%
ENSCAFG00000004436	RB1	94.56	93.32	91.83	91.66	90.83	94.34	100 GENES MIDDLE 40%
ENSCAFG00000001618	CCND3	96.00	93.84	91.78	94.52	93.81	97.35	100 GENES MIDDLE 40%
ENSCAFG00000006162	SMAD9	92.29	89.29	91.68	96.74	96.30	95.74	100 GENES MIDDLE 40%
ENSCAFG00000001844	BAG1	88.66	87.00	91.24	71.00	71.00	86.19	100 GENES MIDDLE 40%
ENSCAFG00000017550	S100A4	95.56	93.00	91.11	93.00	89.00	95.05	100 GENES MIDDLE 40%
ENSCAFG00000009583	EEF1E1	92.26	87.36	90.75	89.03	87.28	95.95	100 GENES MIDDLE 40%
ENSCAFG00000000671	SOD2	91.44	92.00	90.54	90.00	90.00	91.44	100 GENES MIDDLE 40%
ENSCAFG00000010627	RASSF1	88.00	88.53	90.12	90.99	92.00	94.77	100 GENES MIDDLE 40%
ENSCAFG00000017757	PPM1D	90.91	85.41	90.08	88.26	86.94	94.38	100 GENES MIDDLE 40%
ENSCAFG00000007338	NR4A1	89.82	83.03	89.46	88.19	88.69	95.82	100 GENES MIDDLE 40%

ENSCAFG00000007985	PLCE1	91.00	87.18	89.15	84.32	88.23	93.05	100 GENES MIDDLE 40%
ENSCAFG00000007173	NPM1	96.45	93.00	89.05	93.00	87.00	90.88	100 GENES MIDDLE 40%
ENSCAFG00000006142	DCN	87.33	78.89	88.89	80.50	80.83	92.80	100 GENES MIDDLE 40%
ENSCAFG00000004024	RINT1	88.19	85.30	88.85	87.63	86.11	93.06	100 GENES MIDDLE 40%
ENSCAFG00000000867	RAD50	88.19	85.30	88.85	87.63	86.11	93.06	100 GENES MIDDLE 40%
ENSCAFG00000004931	PLCD1	89.99	87.65	88.70	90.87	89.81	93.82	100 GENES MIDDLE 40%
ENSCAFG00000019438	TSC2	89.33	88.99	88.60	91.23	89.22	90.44	100 GENES MIDDLE 40%
ENSCAFG00000013762	PTGS2	87.39	87.25	88.40	87.12	90.78	90.78	100 GENES MIDDLE 40%
ENSCAFG00000019538	STK11	89.41	88.99	88.37	89.50	89.12	91.38	100 GENES MIDDLE 40%
ENSCAFG00000011966	PSME2	87.69	85.00	87.89	94.00	88.00	91.57	100 GENES MIDDLE 40%
ENSCAFG00000004989	ACOX1	88.07	88.07	87.61	88.05	88.35	93.19	100 GENES MIDDLE 40%
ENSCAFG00000016705	TBX21	86.65	86.18	87.56	86.19	87.48	93.27	100 GENES MIDDLE 40%
ENSCAFG00000011816	DGKD	88.72	87.88	87.44	93.25	91.97	93.06	100 GENES MIDDLE 40%
ENSCAFG00000014117	MMP11	91.67	85.40	87.17	88.48	83.58	86.75	100 GENES MIDDLE 40%
ENSCAFG00000018916	SOCS1	96.21	87.74	87.10	87.74	78.30	88.71	100 GENES MIDDLE 40%
ENSCAFG00000017207	CCNG1	87.46	84.07	86.90	92.52	91.86	97.64	100 GENES MIDDLE 40%
ENSCAFG00000019919	IRF8	88.97	86.38	86.85	89.67	88.50	91.78	100 GENES MIDDLE 40%
ENSCAFG00000020252	NQO1	87.91	84.00	86.72	86.00	85.00	88.19	100 GENES MIDDLE 40%
ENSCAFG00000013350	SIPA1	89.84	89.06	86.65	90.11	83.35	88.56	100 GENES MIDDLE 40%
ENSCAFG00000019882	TSC1	88.69	86.00	86.63	86.98	83.75	88.52	100 GENES MIDDLE 40%
ENSCAFG00000000280	CDK4	97.69	95.38	86.49	94.72	94.06	97.03	100 GENES MIDDLE 40%
ENSCAFG00000002065	KIT	89.16	85.60	86.48	82.53	81.87	87.29	100 GENES MIDDLE 40%
ENSCAFG00000003374	IKZF1	88.76	87.18	86.39	92.68	89.21	91.33	100 GENES MIDDLE 40%
ENSCAFG00000007435	E2F1	84.45	82.00	86.29	84.11	85.54	91.20	100 GENES MIDDLE 40%
ENSCAFG00000009802	CYLD	96.36	95.93	86.23	94.77	94.87	97.17	100 GENES MIDDLE 40%
ENSCAFG00000004694	PTCH2	90.38	88.75	86.00	91.02	85.74	88.94	100 GENES MIDDLE 40%
ENSCAFG00000006091	KITLG	83.96	86.00	86.00	83.00	81.00	75.00	100 GENES MIDDLE 40%
ENSCAFG00000001043	XRCC6	84.02	83.88	85.86	83.00	84.00	84.05	100 GENES MIDDLE 40%
ENSCAFG00000009107	NKX3-1	85.71	82.01	85.71	63.08	80.85	84.29	100 GENES MIDDLE 40%
ENSCAFG00000001246	PTCH1	94.00	94.28	85.57	95.73	94.66	95.80	100 GENES MIDDLE 40%
ENSCAFG00000010032	HPGDS	84.42	84.38	85.43	80.00	80.00	87.44	100 GENES MIDDLE 40%
ENSCAFG00000016160	EPHX1	83.74	82.86	85.43	83.52	82.86	84.99	100 GENES MIDDLE 40%
ENSCAFG00000018638	KSR1	85.42	83.48	85.05	83.93	87.44	91.74	100 GENES MIDDLE 40%
ENSCAFG00000031443	EREG	86.96	84.00	84.78	81.00	80.00	89.19	100 GENES MIDDLE 40%
ENSCAFG00000018619	PTGER4	87.76	88.93	84.55	86.15	85.98	87.47	100 GENES MIDDLE 40%
ENSCAFG00000011443	IL10	82.29	74.00	84.52	73.00	73.00	82.69	100 GENES MIDDLE 40%

ENSCAFG00000019249	RPA1	79.53	80.64	84.40	83.12	84.41	88.47	100 GENES MIDDLE 40%
ENSCAFG00000023647	KRT19	82.08	80.13	84.36	84.04	87.42	90.80	100 GENES MIDDLE 40%
ENSCAFG00000016925	FOXO4	87.15	87.75	84.02	88.34	85.40	90.73	100 GENES MIDDLE 40%
ENSCAFG00000009418	TP53INP1	89.00	82.16	83.82	87.50	83.75	87.92	100 GENES MIDDLE 40%
ENSCAFG00000017567	SMAD2	82.00	89.70	83.58	99.57	99.15	99.57	100 GENES MIDDLE 40%
ENSCAFG00000019605	IKBKG	83.00	79.27	83.47	86.52	85.38	89.83	100 GENES MIDDLE 40%
ENSCAFG00000007122	HCK	84.42	84.00	83.27	89.90	88.57	91.63	100 GENES MIDDLE 40%
ENSCAFG00000020375	E2F4	89.70	86.18	83.25	91.79	83.83	85.82	100 GENES MIDDLE 40%
ENSCAFG00000007046	MOS	76.45	78.43	83.14	73.00	70.00	83.00	100 GENES MIDDLE 40%
ENSCAFG00000011924	PSME1	83.13	82.00	83.13	95.00	96.00	98.80	100 GENES MIDDLE 40%
ENSCAFG00000028752	OVCA2	85.02	85.90	82.89	83.26	80.70	82.89	100 GENES MIDDLE 40%
ENSCAFG00000008349	DOK1	83.23	81.34	82.81	83.44	83.23	87.99	100 GENES MIDDLE 40%
ENSCAFG00000000074	MMP19	87.29	78.20	82.23	81.21	77.84	84.83	100 GENES MIDDLE 40%
ENSCAFG00000004964	XRCC2	82.14	76.00	82.14	78.00	76.00	82.86	100 GENES MIDDLE 40%
ENSCAFG00000016310	IKZF3	77.00	74.51	82.09	86.64	88.26	94.81	100 GENES MIDDLE 40%
ENSCAFG00000001910	POLH	85.69	77.50	82.07	78.61	76.22	78.50	100 GENES MIDDLE 40%
ENSCAFG00000011349	GSTP1	80.48	82.00	81.73	85.00	87.00	86.67	100 GENES MIDDLE 40%
ENSCAFG00000018642	NOS2	82.53	79.83	81.66	80.78	79.91	87.15	100 GENES MIDDLE 40%
ENSCAFG00000005808	MYF5	92.55	88.24	81.51	89.41	77.74	84.53	100 GENES MIDDLE 40%
ENSCAFG00000016016	ACADVL	82.00	100.00	81.30	86.14	83.74	90.88	100 GENES MIDDLE 40%
ENSCAFG00000009822	HTATIP2	88.71	91.00	81.10	85.00	63.00	86.83	100 GENES MIDDLE 40%
ENSCAFG00000011231	LYST	86.00	81.00	81.02	85.32	83.95	90.82	100 GENES MIDDLE 40%
ENSCAFG00000004162	MRE11A	84.00	77.49	81.00	87.15	86.38	92.00	100 GENES MIDDLE 40%
ENSCAFG00000015653	INHA	82.20	81.00	80.98	80.00	80.00	79.44	100 GENES MIDDLE 40%
ENSCAFG00000009258	CYTIP	83.57	79.00	80.89	81.00	79.00	84.49	100 GENES MIDDLE 40%
ENSCAFG00000029284	SAT1	97.08	98.00	80.79	97.00	81.00	80.13	100 GENES MIDDLE 40%
ENSCAFG00000007076	RET	82.29	78.64	80.47	83.24	83.24	87.42	100 GENES MIDDLE 40%
ENSCAFG00000009314	NQO2	80.11	82.00	80.46	82.00	80.00	81.20	100 GENES MIDDLE 40%
ENSCAFG00000016463	MAD1L1	84.59	84.11	80.38	81.00	80.00	80.83	100 GENES MIDDLE 40%
ENSCAFG00000015758	EXO1	80.82	74.05	80.02	72.94	72.18	81.06	100 GENES MIDDLE 40%
ENSCAFG00000016404	TP53BP2	78.57	75.58	80.00	80.24	80.07	88.00	100 GENES MIDDLE 40%
ENSCAFG00000017892	DNMT1	81.00	73.53	79.90	77.41	76.43	89.86	100 GENES MIDDLE 40%
ENSCAFG00000008852	DDB2	66.98	78.12	79.87	78.45	81.04	86.85	100 GENES MIDDLE 40%
ENSCAFG00000016714	TP53	83.21	78.01	79.85	77.35	72.89	77.92	100 GENES MIDDLE 40%
ENSCAFG00000012508	TP53BP1	83.21	78.01	79.85	77.35	72.89	77.92	100 GENES MIDDLE 40%
ENSCAFG00000004606	BIN1	95.00	93.31	79.74	94.77	78.58	95.24	100 GENES MIDDLE 40%

ENSCAFG00000000426	LYZ	82.43	77.03	79.73	76.00	82.00	80.41	100 GENES MIDDLE 40%
ENSCAFG00000018216	RBBP8	80.00	76.34	79.54	76.03	75.38	81.56	100 GENES MIDDLE 40%
ENSCAFG00000013263	MGMT	85.39	85.00	79.31	69.00	66.00	65.31	100 GENES MIDDLE 40%
ENSCAFG00000000120	IL23A	87.37	70.90	79.07	75.00	69.00	79.17	100 GENES MIDDLE 40%
ENSCAFG00000009293	IQGAP2	81.33	77.43	79.07	88.95	85.40	89.75	100 GENES MIDDLE 40%
ENSCAFG00000008434	ATP2A2	97.00	79.00	79.00	98.38	98.21	99.00	100 GENES MIDDLE 40%
ENSCAFG00000024739	TRIM24	85.49	90.00	79.00	93.00	92.00	97.00	100 GENES MIDDLE 40%
ENSCAFG00000019676	ERRFI1	82.94	77.92	78.96	81.86	78.52	84.42	100 GENES MIDDLE 40%
ENSCAFG00000011919	CHEK2	75.00	77.00	78.91	86.07	86.16	86.00	100 GENES MIDDLE 40%
ENSCAFG00000003465	EGFR	89.92	88.66	78.90	90.35	90.81	92.31	100 GENES MIDDLE 40%
ENSCAFG00000006934	AMHR2	85.00	81.88	78.83	77.74	76.74	81.64	100 GENES MIDDLE 40%
ENSCAFG00000002448	AHR	77.54	69.07	78.81	70.19	69.85	84.14	78 genes BOTTOM 30%
ENSCAFG00000020305	CDH1	82.00	80.60	78.74	81.23	80.31	81.58	78 genes BOTTOM 30%
ENSCAFG00000005182	FANCD2	78.21	72.30	78.64	74.93	75.82	84.15	78 genes BOTTOM 30%
ENSCAFG00000017146	CXCR3	80.00	76.38	78.51	86.49	83.77	89.34	78 genes BOTTOM 30%
ENSCAFG00000028905	CDKN2C	94.64	93.45	78.51	92.26	89.88	95.24	78 genes BOTTOM 30%
ENSCAFG00000009209	NR4A2	77.00	96.17	78.26	91.73	86.53	86.77	78 genes BOTTOM 30%
ENSCAFG00000010700	CCND1	96.61	95.59	78.08	93.90	88.58	92.69	78 genes BOTTOM 30%
ENSCAFG00000011930	CCNDBP1	81.85	76.00	78.00	80.69	77.26	84.55	78 genes BOTTOM 30%
ENSCAFG00000018687	MLLT1	94.68	87.32	77.74	82.77	72.93	76.23	78 genes BOTTOM 30%
ENSCAFG00000019526	FUT7	86.32	82.59	77.45	80.00	71.00	81.05	78 genes BOTTOM 30%
ENSCAFG00000015177	SELL	81.64	75.00	76.94	76.08	73.85	80.53	78 genes BOTTOM 30%
ENSCAFG00000011555	MFGE8	64.56	78.00	76.60	57.00	72.00	65.03	78 genes BOTTOM 30%
ENSCAFG00000000150	POLI	79.60	71.31	76.24	77.45	74.44	87.18	78 genes BOTTOM 30%
ENSCAFG00000013217	GPRC5A	76.97	78.93	76.19	76.14	75.64	77.59	78 genes BOTTOM 30%
ENSCAFG00000009905	MMP9	83.60	69.93	75.74	81.35	78.43	79.63	78 genes BOTTOM 30%
ENSCAFG00000009680	GCNT2	80.60	76.44	75.62	86.28	83.04	79.35	78 genes BOTTOM 30%
ENSCAFG00000016361	PRDM2	77.63	76.12	75.57	78.78	75.74	98.09	78 genes BOTTOM 30%
ENSCAFG00000009086	IFNAR1	60.63	50.00	75.55	49.00	61.00	65.80	78 genes BOTTOM 30%
ENSCAFG00000025533	ITIH4	60.63	50.00	75.55	49.00	61.00	65.80	78 genes BOTTOM 30%
ENSCAFG00000018538	SERBP1	79.85	74.88	75.50	79.10	75.99	85.64	78 genes BOTTOM 30%
ENSCAFG00000008636	XRCC4	73.90	84.00	75.17	83.22	84.05	79.17	78 genes BOTTOM 30%
ENSCAFG00000015559	PMS2	80.51	77.25	75.00	77.62	72.11	77.05	78 genes BOTTOM 30%
ENSCAFG00000001680	CDKN2B	81.88	84.38	74.55	88.28	82.81	81.16	78 genes BOTTOM 30%
ENSCAFG00000010024	DOK2	74.70	75.96	74.28	74.04	72.36	84.22	78 genes BOTTOM 30%
ENSCAFG00000015190	PLAU	76.32	71.47	74.13	71.22	70.14	78.98	78 genes BOTTOM 30%

ENSCAFG00000006410	WRN	74.88	72.82	74.12	69.18	68.87	74.27	78 genes BOTTOM 30%
ENSCAFG00000015080	MMP7	71.10	68.00	73.95	70.00	66.00	78.66	78 genes BOTTOM 30%
ENSCAFG00000014488	BAD	76.79	99.00	73.53	75.00	70.00	79.88	78 genes BOTTOM 30%
ENSCAFG00000018001	NEIL1	78.71	74.51	73.48	79.49	75.32	78.97	78 genes BOTTOM 30%
ENSCAFG00000014612	CXCR2	72.22	71.30	73.47	71.11	71.88	75.49	78 genes BOTTOM 30%
ENSCAFG00000015642	FOXM1	72.74	72.14	73.04	80.16	80.21	83.01	78 genes BOTTOM 30%
ENSCAFG00000025384	PML	75.00	62.98	72.90	69.34	66.67	80.09	78 genes BOTTOM 30%
ENSCAFG00000004448	ERCC1	73.00	73.97	72.33	86.20	82.49	88.51	78 genes BOTTOM 30%
ENSCAFG00000008279	PTPRJ	75.03	72.25	72.13	71.00	68.82	72.33	78 genes BOTTOM 30%
ENSCAFG00000015425	BHLHA15	73.98	73.10	71.50	72.14	74.62	76.19	78 genes BOTTOM 30%
ENSCAFG00000010055	LZTS1	86.99	70.12	71.43	83.23	79.75	80.84	78 genes BOTTOM 30%
ENSCAFG00000004867	MBD4	75.00	67.50	71.40	66.26	66.44	76.95	78 genes BOTTOM 30%
ENSCAFG00000012385	BLM	76.00	67.55	71.08	75.52	73.06	81.73	78 genes BOTTOM 30%
ENSCAFG00000002797	KLF4	97.71	98.00	71.00	85.80	78.37	79.00	78 genes BOTTOM 30%
ENSCAFG00000019487	TCF3	62.00	65.86	70.95	81.91	78.27	84.27	78 genes BOTTOM 30%
ENSCAFG00000020110	F3	68.47	54.00	70.82	55.00	57.00	73.00	78 genes BOTTOM 30%
ENSCAFG00000015383	TNFSF10	70.43	60.00	70.67	65.00	63.00	78.72	78 genes BOTTOM 30%
ENSCAFG00000016995	PGF	73.85	62.00	70.45	66.00	65.00	85.29	78 genes BOTTOM 30%
ENSCAFG00000032433	PSCA	66.27	66.00	68.67	59.00	63.00	74.70	78 genes BOTTOM 30%
ENSCAFG00000000820	CSF2	70.14	56.94	67.36	54.86	52.78	70.14	78 genes BOTTOM 30%
ENSCAFG00000031861	MZF1	37.00	35.00	67.14	82.74	72.59	75.39	78 genes BOTTOM 30%
ENSCAFG00000013658	NR0B1	70.04	61.68	66.88	65.47	64.62	72.61	78 genes BOTTOM 30%
ENSCAFG00000019533	PTGDS	66.67	63.00	66.86	72.00	69.00	76.33	78 genes BOTTOM 30%
ENSCAFG00000006164	CYP1B1	66.30	84.66	65.50	81.03	79.37	81.95	78 genes BOTTOM 30%
ENSCAFG00000004467	XPC	68.83	66.95	64.95	75.17	68.87	75.18	78 genes BOTTOM 30%
ENSCAFG00000017037	MUC1	52.19	46.00	61.93	55.00	63.19	72.80	78 genes BOTTOM 30%
ENSCAFG00000010221	SUFU	61.95	60.68	61.46	97.73	97.78	98.89	78 genes BOTTOM 30%
ENSCAFG00000000408	IFNG	61.45	37.82	61.45	41.03	44.87	65.66	78 genes BOTTOM 30%
ENSCAFG00000017941	CYP1A2	70.00	61.00	61.00	72.76	71.35	82.00	78 genes BOTTOM 30%
ENSCAFG00000010739	TERT	59.46	55.28	60.98	61.69	59.00	70.49	78 genes BOTTOM 30%
ENSCAFG00000014003	SRPX	57.33	58.71	60.75	58.30	57.12	66.93	78 genes BOTTOM 30%
ENSCAFG00000004008	IL2	65.58	71.00	60.00	73.68	67.44	74.19	78 genes BOTTOM 30%
ENSCAFG00000028557	NKX2-8	87.39	74.77	60.00	74.06	52.92	68.00	78 genes BOTTOM 30%
ENSCAFG00000003833	SDC1	67.94	62.00	59.93	76.00	74.00	79.44	78 genes BOTTOM 30%
ENSCAFG00000004696	SUV39H2	58.00	58.00	58.00	89.75	81.53	97.00	78 genes BOTTOM 30%
ENSCAFG00000010351	TFF1	59.00	58.00	55.00	62.07	55.17	69.00	78 genes BOTTOM 30%

ENSCAFG00000011441	CASC1	62.00	51.00	54.95	65.45	57.86	73.08	78 genes BOTTOM 30%
ENSCAFG00000025377	AHRR	62.00	62.61	54.65	59.47	51.52	58.88	78 genes BOTTOM 30%
ENSCAFG00000028875	CD248	72.79	68.29	54.18	76.61	53.15	71.30	78 genes BOTTOM 30%
ENSCAFG00000014600	BRCA1	55.00	43.16	50.82	55.93	53.37	74.01	78 genes BOTTOM 30%
ENSCAFG00000006383	BRCA2	61.49	48.79	48.10	57.43	53.61	69.18	78 genes BOTTOM 30%
ENSCAFG00000001635	MLLT3	61.49	48.79	48.10	57.43	53.61	69.18	78 genes BOTTOM 30%
ENSCAFG00000002420	MEOX2	44.00	43.00	45.00	96.71	98.35	99.00	78 genes BOTTOM 30%
ENSCAFG00000005133	LTBP4	95.23	38.00	42.00	84.01	91.38	94.00	78 genes BOTTOM 30%
ENSCAFG00000002733	IL6	40.24	43.00	41.33	41.00	40.00	59.91	78 genes BOTTOM 30%
ENSCAFG00000012509	TAF4	83.71	85.96	41.00	94.39	96.25	97.00	78 genes BOTTOM 30%
ENSCAFG00000013874	CXCR6	36.00	39.00	40.00	75.22	80.61	0.00	78 genes BOTTOM 30%
ENSCAFG00000017100	ADAM15	39.00	38.00	40.00	78.60	74.79	80.00	78 genes BOTTOM 30%
ENSCAFG00000023237	SPN	55.00	46.36	40.00	48.25	44.70	45.00	78 genes BOTTOM 30%
ENSCAFG00000002121	PDCD1LG2	38.00	34.00	37.00	71.90	67.16	68.00	78 genes BOTTOM 30%
ENSCAFG00000018946	TICAM1	32.00	31.00	36.00	52.11	50.35	66.00	78 genes BOTTOM 30%
ENSCAFG00000014846	SKIL	31.00	32.00	32.00	88.01	87.17	91.00	78 genes BOTTOM 30%
ENSCAFG00000001819	DMTF1	27.00	27.00	27.00	95.14	93.97	97.00	78 genes BOTTOM 30%

Table 1: 256 Pairwise Reciprocal Best-Hit Tumorigenesis Orthologs Mapped Between Human, Dog, Mouse and Naked Mole Rat.

Upon inspection of the heat map, the strong red visible pattern of sequence identity is observed within the first 24 tumorigenesis orthologs. Correspondingly, these top 24 orthologs represent protein coding genes that are extremely well conserved across all species in the analysis. For example, the most conserved gene, exhibits 100% amino acid identity between all six pairs of species comparisons in the analysis. The second most conserved gene is 100% identical between three pairs of species and at least 99% identical across the remaining species pairs. Within the set of 24 most conserved protein orthologs between dog and naked mole rat, the amino acid identity across all pair-wise species comparisons is greater than 90% identity. Similarly, the pattern of protein identity displayed in the heat map conveys that the least conserved orthologs between dog and naked mole rat correspond to the bottom 24 orthologs, which display mostly white and blue colors corresponding to protein identities as less than 62% identity between dog and naked mole rat and for which 13 tumorigenesis orthologs exhibit less than 50% identity, and the remaining 11 orthologs exhibiting identity between 50.83% and 60.98% identity between dog and naked mole rat.

Based on the distribution of protein sequence identity scores in Table 1 and visualized within the heat map shown in Figure 2, the tumorigenesis orthologs were grouped into three categories based on their ranked placement corresponding to the following groups:

(1) TOP 30% - the top 30% most conserved tumorigenesis orthologs (78 genes) with protein percent identity ranging from 92.21% to 100% between dog and naked mole rat;

(2) MIDDLE 40% - the middle 40% conserved tumorigenesis orthologs (100 genes) with protein percent identity ranging from 78.83% to 92.19% between dog and naked mole rat;

(3) BOTTOM 30% - the bottom least conserved (most divergent) tumorigenesis orthologs (78 genes) with protein identity ranging from 27% to 78.81% between dog and naked mole rat.

The protein percent identity classifications were applied in downstream bioinformatics and comparative genomics analyses to investigate the relationship, if any, that might exist between the extent of protein identity and the functional role of these orthologs in health and disease.

Analysis of human, mouse and dog tumorigenesis orthologs in pubmed abstracts

In order to gain a better understanding of how tumorigenesis orthologs are represented across human, dog and mouse publications indexed in the PubMed database, species specific PubMed cancer-related queries were generated and automatically executed for each ortholog within each species using PubMatrix. The results provided information on the distribution of published papers across the orthologs and species. From among the 256 protein orthologs, 255 were associated with abstracts associated with human cancer papers, while 253 were associated with abstracts in association with mouse cancer publications and just 96 of the tumorigenesis ortholgs were associated with an abstract in the context of a dog cancer publication. Interestingly, all 96 tumorigenesis orthologs associated with dog cancer publications were included in the 253 orthologs associated with mouse cancer publications, and the entire set of mouse published orthologs

were included within the set of 255 human tumorigenesis publications (Figure 3A).

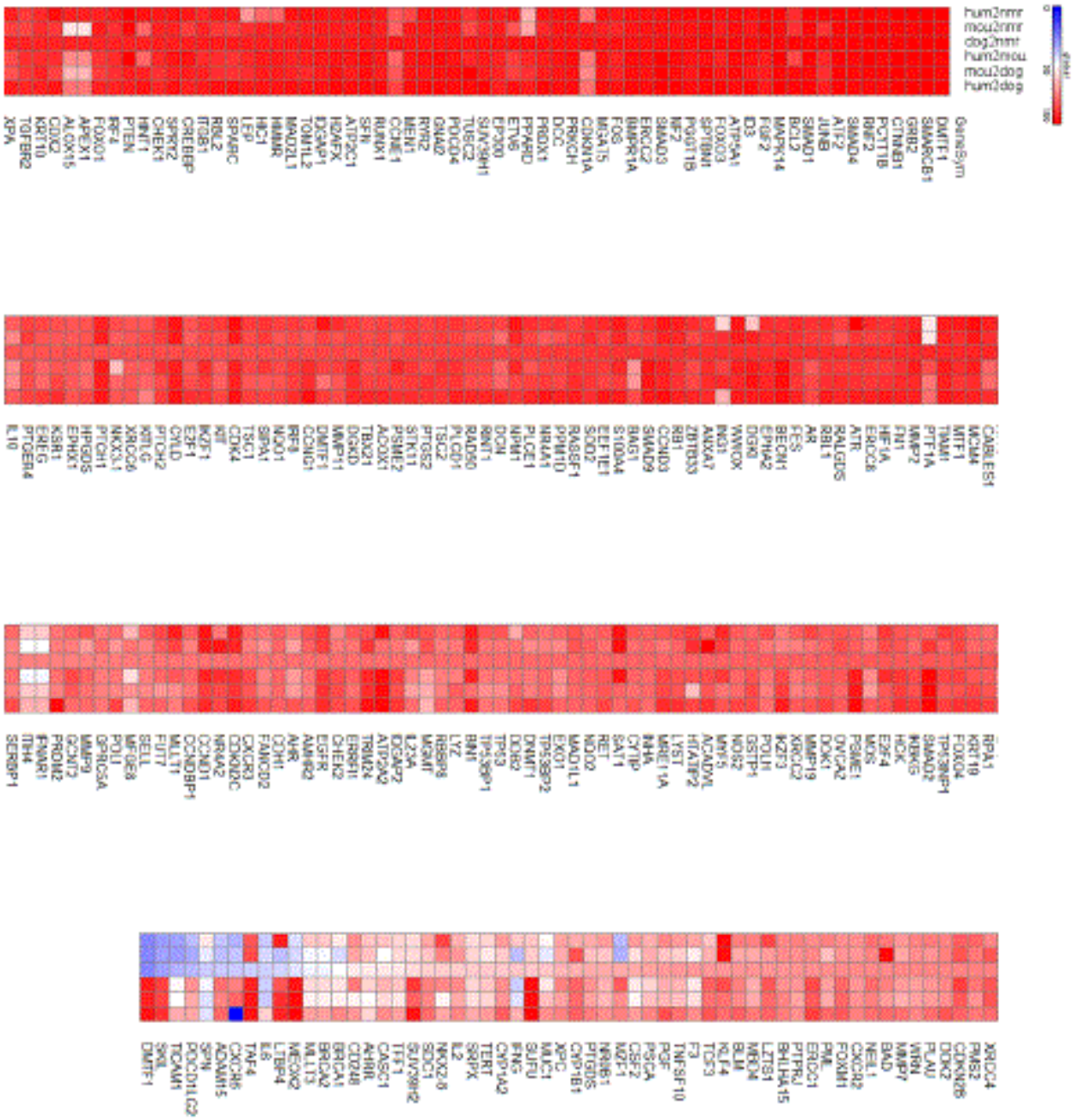

Figure 2: Heat Map of Sequence Identity among Tumorigenesis Orthologs. Visualization of the protein pairwise percent identity between each pair of species was achieved using a heat map in which each of the tumorigenesis orthologs was represented by a column and each species pair-wise comparison was represented by a row. Individual cells in the heat map correspond to the percent identity for an individual ortholog between two species. Dark red represents the highest percent identity, while white represents moderate percent identity and blue represents the lowest percent identity between orthologous protein sequences between the species. Note that the heat map is ordered by decreasing percent identity between dog and naked mole rat orthologs.

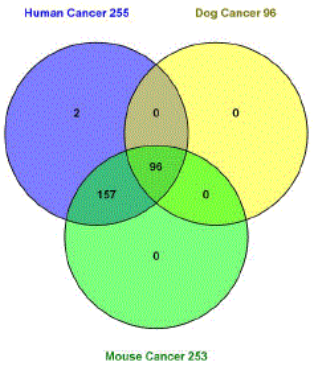

Figure 3A: Venn Diagram of Tumorigenesis Orthologs Published in Each Species. The number of orthologs within each species for which at least a single pubmed abstract was identified connecting the ortholog to the species (i.e., human, dog, mouse) and cancer. Out of 256 total orthologs, 255 human orthologs were associated with pubmed publications. Similarly, 253 mouse orthologs were associated with pubmed abstracts while just 96 dog orthologs were published in the context of cancer.

Next, the same species cancer relationships were investigated in PubMed to assess the extent of tumorigenesis orthologs that that were associated with publication abstracts relating to single nucleotide polymorphisms (SNPs). Once again, PubMatrix was used to automate the query generation and query execution portions of the analysis. The results indicated that, within human cancer SNP publications, 209 tumorigenesis orthologs were associated with at least one abstract while mouse and dog were associated with 106 and 15 tumorigenesis ortholog polymorphism cancer publications, respectively (Figure 3B).

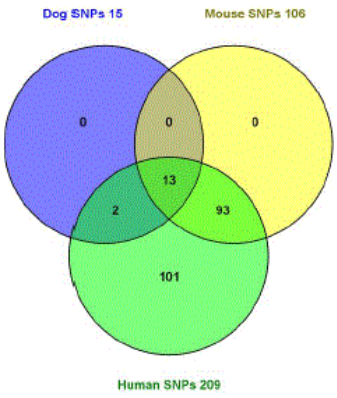

Figure 3B: Venn Diagram of Tumorigenesis SNPs Published in Each Species. The number of orthologs within each species for which at least a single pubmed abstract was identified connecting the ortholog to the species (i.e., human, dog, mouse) and a SNP and cancer were identified. Out of 256 total orthologs, 209 human orthologs were associated with SNPs in pubmed publications. In contrast, only 106 mouse orthologs were associated with polymorphisms and pubmed abstracts while just 15 dog orthologs were published in the context of cancer and genetic variation.

Because fewer dog tumorigenesis orthologs (96) are represented by published papers in the PubMed database compared to human (255) and mouse (253) orthologs it was worthwhile to investigate what, if any, relationship might exist between the extent of protein identity and the likelihood of observing at least one associated dog cancer- related abstract. The rationale for this line of reasoning was based on the possibility that dog tumorigenesis publications might be biased towards orthologs exhibiting higher levels of protein identity, as might be expected to occur, if, for example, canine cancer gene studies were based on sequence conservation to human oncogenes and tumor suppressors.

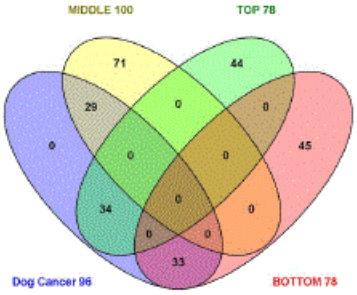

Figure 4A: Published Dog Tumorigenesis Orthologs Mapped to Percent Identity. The number of dog orthologs associated with at least one pubmed cancer abstract is displayed by ortholog conservation and divergence. Almost the same number of dog orthologs were associated with published abstracts within the TOP 30% (34 orthologs) and the BOTTOM 30% (33 orthologs). These results provide evidence that the percent identity is independent of whether the ortholog has been published in the context of cancer.

Among the set of 96 dog tumorigenesis orthologs associated with at least one PubMed cancer related abstract, 33 orthologs were associated with the set of BOTTOM 30% least conserved orthologs while 34 orthologs were part of the TOP 30% most conserved and the remaining 29 orthologs associated with PubMed cancer publications were located with the MIDDLE 40% of the tumorigenesis orthologs. The results demonstrate that the representation of most conserved (34 orthologs) AND least conserved (33 orthologs) is almost identical (Figure 4A). Similar representations of ortholog identity were observed within the mouse data with 77 most conserved mouse orthologs associated with cancer abstracts, 78 least conserved mouse orthologs, and the remaining 98 mouse orthologs that were associated with cancer abstracts were derived from the MIDDLE 40% percent identity (Figure 4B). Finally, an almost identical distribution of tumorigenesis orthologs was observed for human orthologs associated with cancer abstracts, with 77 orthologs from the TOP 30% most conserved, 78 orthologs from the BOTTOM 30% least conserved and 100 from MIDDLE 40% conserved (Figure 4C).

Gene Ontology Enrichment of Conserved and Divergent Tumorigenesis Orthologs Gene ontology (GO) enrichment analysis was performed to identify GO annotations enriched within the TOP 30% most conserved tumorigenesis orthologs and within the least conserved tumorgenesis orthologs. Many statistically significant GO terms enriched in the TOP 30% most conserved orthologs are associated with embryogenesis, development, gastrulation, cell differentiation, germ cell development and organogenesis. Specifically, the following GO terms with p-values less than 10-3 were identified:

reproductive process; developmental process involved in reproduction, germ cell development, anatomical structure development, organ morphogenesis, branching involved in ureteric bud morphogenesis, odontogenesis of dentin containing tooth, gastrulation with mouth forming second, cell fate commitment, cell differentiation, epithelial cell differentiation, anterior/posterior axis specification, muscle cell proliferation, striated muscle cell proliferation, and cardiac muscle proliferation (Figure 5A and 5B).

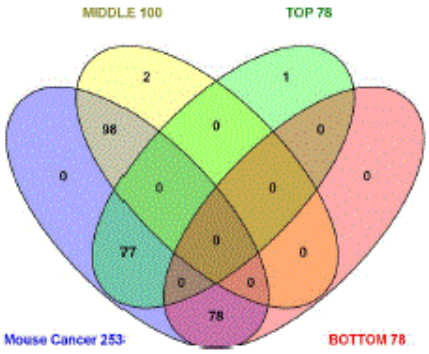

Figure 4B: Published Mouse Tumorigenesis Orthologs Mapped to Percent Identity. The number of mouse orthologs associated with at least one pubmed cancer abstract is displayed by ortholog conservation and divergence. Almost the same number of mouse orthologs were associated with published abstracts within the TOP 30% (78 orthologs) and the BOTTOM 30% (77 orthologs). These results provide evidence that the percent identity is independent of whether the ortholog has been published in the context of cancer.

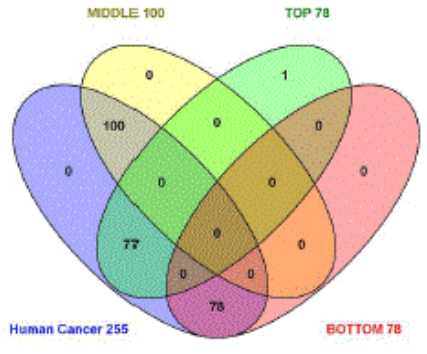

Figure 4C: Published Human Tumorigenesis Orthologs Mapped to Percent Identity. The number of human orthologs associated with at least one pubmed cancer abstract is displayed by ortholog conservation and divergence. Almost the same number of human orthologs were associated with published abstracts within the TOP 30% (78 orthologs) and the BOTTOM 30% (77 orthologs). These results provide evidence that the percent identity is independent of whether the ortholog has been published in the context of cancer. Note that the pattern is very similar to that identified for mouse, except that the human MIDDLE orthologs have 100 associated cancer publications while the mouse has 98 associated with cancer publications.

Additional GO biological process terms enriched within these most conserved tumorigenesis orthologs are regulation of heart morphogenesis, regulation of cell proliferation involved in heart morphogenesis, regulation of muscle tissue development, regulation of muscle organ development, positive regulation of ossification, positive regulation of cell differentiation, and positive regulation of osteoblast differentiation (Figure 5C). In contrast to the strong developmental themes identified in the most conserved tumorigenesis orthologs, the least conserved orthologs were enriched for four very specific GO terms with p-values less than 10-3 associated with immune function: positive regulation of leukocyte proliferation, positive regulation of mononuclear cell proliferation, positive regulation of lymphoid proliferation, and positive regulation of T-cell proliferation (Figure 5D).

Phenotype enrichment of conserved and divergent tumorigenesis orthologs

1. The gene ontology enrichment analysis successfully identified some distinct biological processes differentially associated with the most conserved versus the least conserved tumorigenesis orthologs. Therefore, it seemed reasonable that additional functional analysis might further elucidate important roles of these proteins within these two tumor-associated gene sets. Subsequently phenotype enrichment analysis was carried out to identify phenotypes exhibiting statistically significant enrichment within the TOP 30% most conserved tumorigenesis orthologs as well as within the the BOTTOM 30% least conserved tumorigenesis orthologs. The phenotype annotations are organized by broad categories such as body system. Within each broad phenotype category, additional phenotypes are classified more specifically. Each level of the phenotype hierarchical structural organization is assigned a number, which increases as the depth of phenotypes increases. As an example, the following phenotypes are part of the phenotype ontology organizational structure and each phenotype in the structure is associated with its corresponding level (note that more than one phenotype may be associated with a specific level): behavioral/neurological phenotype (level-2), abnormal behavior (level-3), abnormal spatial learning (level 4), enhanced spatial learning (level 5) impaired spatial learning (level 5). The phenotype enrichment analysis employed in this approach utilized phenotype levels 2 through 5, which provided sufficient phenotypic diversity to be biologically informative while simultaneously excluding a very large number of phenotypes which could make the data produced by the analysis overwhelming and ultimately reduce the value of the analysis.

Statistically significant phenotype enriched within the TOP 30% most conserved tumorigenesis orthologs (Table 2) can be classified into five broad categories:

Tumor related phenotypes: tumorigenesis (p-value=1.391E-74), altered tumor susceptibility (p-value=2.817E-67), abnormal tumor incidence (p-value=3.499E-57), altered tumor pathology (p-value=8.196E-26), altered metastatic potential (p- value=1.697E-15), increased skin tumor incidence (p-value=7.541E-07), gastrointestinal tract polyps (p-value=0.00001647), abnormal cell proliferation (p-value=4.151E-08).

Embryogenesis, growth and lethality related phenotypes: embryogenesis phenotype (p-value=1.192E-09), abnormal prenatal

growth/weight/body size (p-value=8.835E-10), prenatal lethality (p-value=5.708E-07), embryonic growth retardation (p-value=2.865E-07), premature death (p-value=1.933E-13).

Developmental related phenotypes: Abnormal vascular development (p-value=1.166E-06), abnormal skeleton development (p-value=7.468E-06), abnormal cardiovascular development (p-value=1.516E-08), abnormal craniofacial development (p-value=0.00006428), abnormal blood cell morphology/development (p-value=2.949E-08), abnormal nervous system development (p-value=0.00002332).

Anatomical morphology related phenotypes: abnormal reproductive system morphology (p value=4.009E-07), abnormal female reproductive system morphology (p- value=5.972E 06), abnormal digestive system morphology (p-value=1.805E-07), abnormal epiphyseal plate morphology (p-value=0.0001711), abnormal pericardium morphology (p-value=0.0005925), abnormal epidermal layer morphology (p- value=0.00001226), abnormal spleen morphology (p-value=0.00001254), abnormal liver morphology (p-value=0.00004149), abnormal hair follicle morphology (p-value=5.266E-06), abnormal extraembryonic tissue morphology (p-value=6.806E-08), abnormal coat/hair morphology (p-value=0.0002076);

Immune, inflammatory, and hematopoietic related phenotypes: immune system phenotype (p-value=6.307E-09), abnormal immune system morphology (p- value=8.912E-07), abnormal immune system physiology (p-value=3.687E-08), abnormal hematopoietic system morphology/development (p-value=1.103E-08), abnormal inflammatory response (p-value=4.211E-07), abnormal blood cell morphology/development (p-value=2.949E-08), abnormal bone marrow cell morphology/development (p-value-=0.00009508), increased inflammatory response (p- value=9.052E-07), abnormal hematopoiesis (p-value=6.487E-08);

Statistically significant phenotypes enriched within the BOTTOM 30% least conserved tumorigenesis orthologs (Table 3) can be classified into three broad categories:

Tumor related phenotypes: tumorigenesis (p-value=5.28E-72), altered tumor susceptibility (p-value=3.85E-61), altered tumor pathology (p-value=1.84E-16), abnormal tumor incidence (p-value=3.5E-57), altered metastatic potential (p- value=2.68E-06), altered tumor morphology (p-value=1.63E-12), increased tumor incidence (p-value=2.72E-46), decreased tumor incidence (7.14E-10), decreased tumor growth/size (8.45E-06), increased incidence of chemically-induced tumors (p-value=2.84E-08), decreased incidence of chemically-induced tumors (p-value=1.31E-05);

Immune, inflammatory, and hematopoietic related phenotypes: immune system phenotype (p-value=2E-08), hematopoietic system phenotype (p-value=3.9E-05), abnormal hematopoietic system morphology/development (p-value=3.78E-05), abnormal immune system morphology (p-value=4.06E-06), abnormal immune system physiology (p-value=6.37E-07), abnormal immune system cell morphology (p- value=1.7E-05), abnormal immune system organ morphology (p-value=3.23E-05), abnormal immune cell physiology (p-value=3.22E-05), abnormal adaptive immunity (p-value=3.37E-05), abnormal cell-mediated immunity (p-value=9.63E-06), abnormal antigen presenting cell physiology (p-value=5.05E-05), abnormal response to infection (p-value=3.73E-05), abnormal leukocyte physiology (p-value=6.67E-05), abnormal antigen presenting cell physiology (p-value=0.000162);

Morphology related phenotypes: morphology/development (p-value=3.78E-05), abnormal spleen morphology (p-value=0.000917), abnormal mammary gland morphology (p-value=5.24E-05), abnormal lymph organ size (p-value=8.58E-06), mammary gland hyperplasia (p-value=0.002), abnormal reproductive system morphology (p-value=0.003), abnormal digestive system physiology (p-value=0.003).

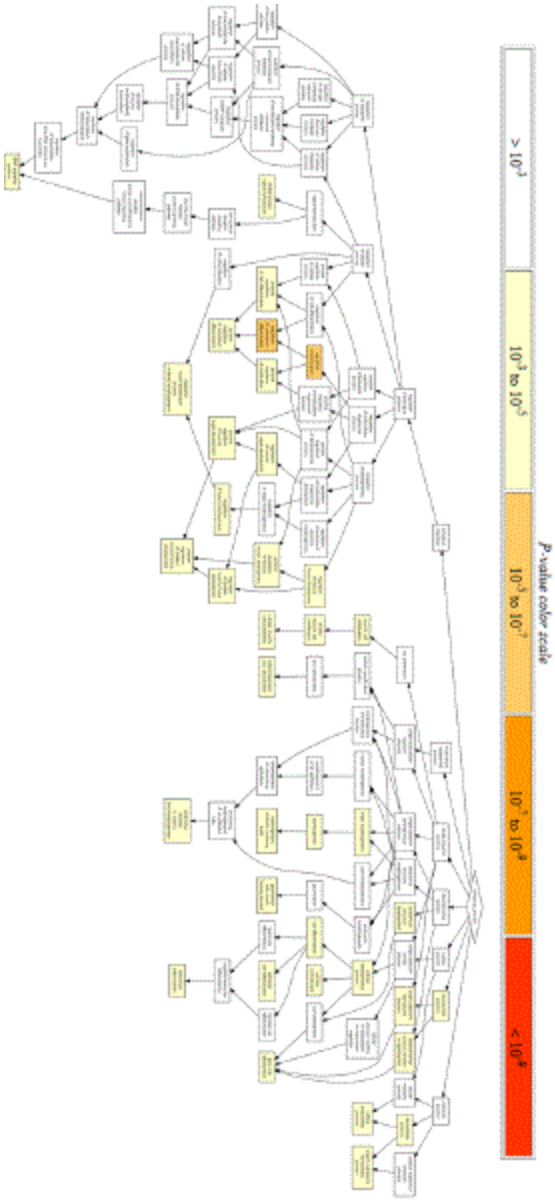

Figure 5A: Gene Ontology Enrichment Top 30% Most Conserved Orthologs (Global). Gene ontology (GO) enrichment analysis was performed to identify GO annotations enriched within the TOP 30% most conserved tumorigenesis orthologs. This figure provides a global view of the gene ontology annotations enriched.

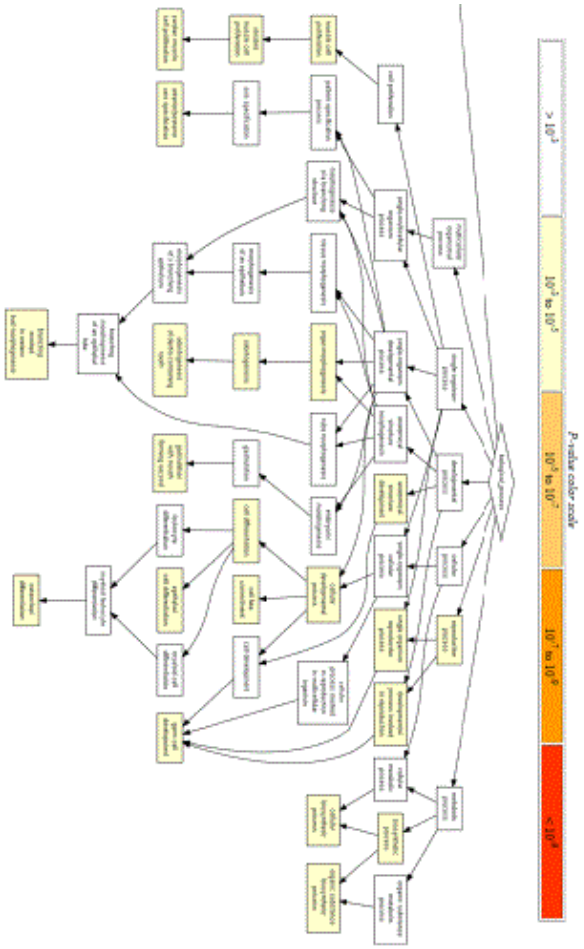

Figure 5B: Zoomed in View of Right Portion of GO Enrichment for Top 30% Orthologs. Gene ontology (GO) enrichment analysis was performed to identify GO annotations enriched within the TOP 30% most conserved tumorigenesis orthologs. This figure provides a view of the gene ontology annotations enriched in the right portion of the gene ontology relationship tree (shown in the global view). Reproductive process; developmental process involved in reproduction, germ cell development, anatomical structure development, organ morphogenesis were statistically significant.

Term co-occurrence analysis and heat map visualization of biological relationships

Having identified functional annotation terms that exhibited statistically significant enrichment within the conserved and/or the divergent tumorigenesis ortholog gene- sets, literature mapping could be employed to identify meaningful associations between the terms. Because many different enriched terms mapped to a core set of biological themes, it seemed plausible that biologically relevant associations between these themes might be uncovered through query expansion gained by leveraging all of the enriched terms for literature mining. This approach enhances query precision and query recall and ultimately maximizes the possibility of identifying subtle relationships between the phenotypes/biological processes.

A representative set of enriched gene ontology terms were selected to query PubMed via the literature blasting interface, PubAtlas. Similarly, representative mammalian phenotype terms were chosen for query PubMed. The selected terms were chosen to reflect the identified themes (for example embryogenesis, tumorigenesis, anatomical morphology, organogenesis, immune system). Literature mapping provided a mechanism for exploring the connectivity among the enriched terms. The results of the gene ontology literature mapping are illustrated in the heat map shown in Figure 6A. The heat map displays the strongest associations between co-occurring terms in red, while weaker associations are represented by a continuum of color ranging from orange (strong associations), to yellow (moderate associations) and ultimately white (no associations). To aide in the visualization of connections between the terms, rectangular outlines were placed on the heat map and the corresponding terms were colored with the same color as the rectangle outlining the heat map colored pixels.

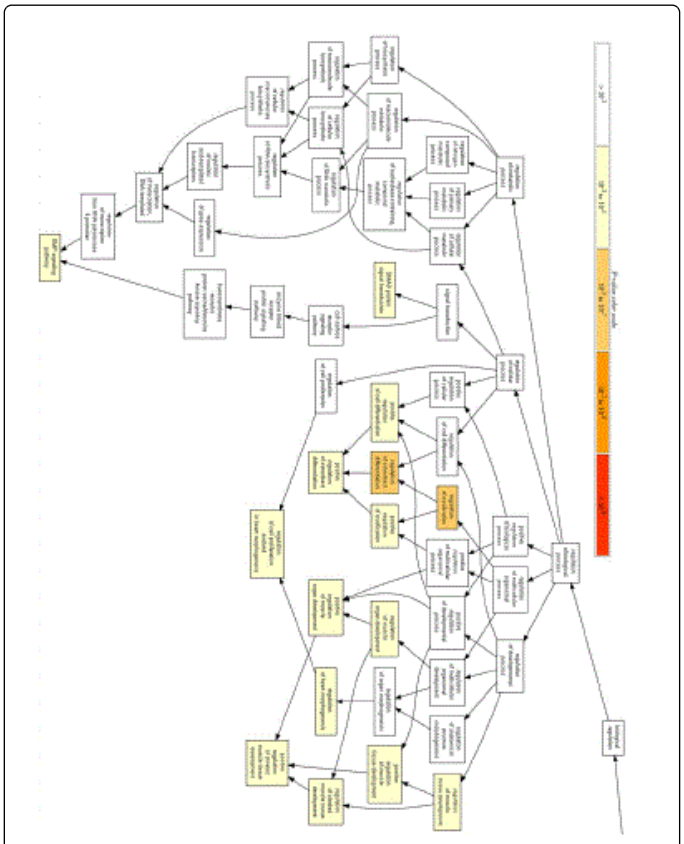

Figure 5C: Zoomed in View of Left Portion of GO Enrichment for Top 30% Orthologs. Gene ontology (GO) enrichment analysis was performed to identify GO annotations enriched within the TOP 30% most conserved tumorigenesis orthologs. This figure provides a view of the gene ontology annotations enriched in the left portion of the gene ontology relationship tree (shown in the global view). Heart morphogenesis, regulation of cell proliferation involved in heart morphogenesis, regulation of muscle tissue development, regulation of muscle organ development, positive regulation of ossification, positive regulation of cell differentiation, and positive regulation of osteoblast differentiation were statistically significant.

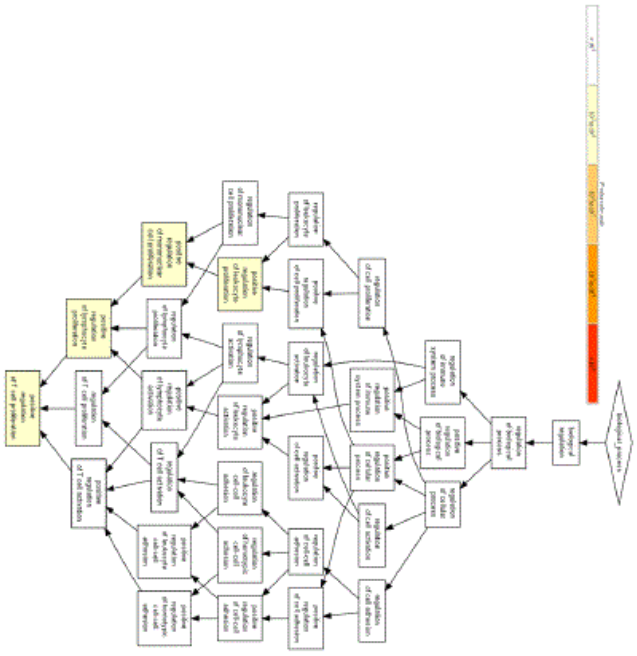

Figure 5D: Gene Ontology Enrichment for Bottom 30% Most Divergent Orthologs. The least conserved orthologs were enriched for four very specific GO terms with p- values less than 10-3 associated with immune function: positive regulation of leukocyte proliferation, positive regulation of mononuclear cell proliferation, positive regulation of lymphoid proliferation, and positive regulation of T-cell proliferation.

As an example, the small green rectangle outlines six orange pixels in the heat map near the upper right hand corner of the heat map shown in Figure 6A. The terms on the right side, aligning to the green square, are also highlighted in green (cell cycle progression, transforming growth factor, TGF beta). The terms on the bottom of the heat map (near the right hand corner) are also highlighted in green because they are also aligned with the green rectangle (T-cell proliferation, lymphocyte proliferation, leukocyte proliferation). This relationship demonstrates a functional biological connection between the cell cycle progression, TFG beta and lymphocyte proliferation. Other connections are represented by additional rectangles and colors. These connections provide biological links between related biological processes.

Within the set of enriched gene ontology terms, the heat map provides examples of connections between different categories of terms. One example is the connection identified between the immune category (monocytes, macrophages) and the development category (osteoclast differentiation, ossification, osteoblast differentiation). Interestingly, the blue cancer terms on the right side of the heat map (osteosarcoma, glioma, hepatocellular carcinoma, pancreatic cancer, on small cell lung cancer, small cell lung cancer) are adjacent to the terms (T cell receptor signaling, T cell signaling), all of which are highlighted in the same shade of blue. The complimentary terms on the side of the heat map (cell cycle progression, transforming growth factor beta, p53 signaling, mTOR, mTOR signaling pathway) further elucidate signaling components underlying tumorigenesis that are important for skeleton development and morphology (the regulation of osteoclast and osteoblast differentiation) as well as immune system function (proliferation of lymphocytes, leukocytes and T-cells).

Connections between enriched mammalian phenotype terms are visualized on the heat map shown in Figure 6B. The light-blue highlighted phenotypes on the right side of the heat map (skeletal morphology, somite development, abnormal skeleton, skeleton development, abnormal facial development, facial morphology) are linked with the light-blue phenotypes on the bottom left corner of the heat (axial skeleton, abnormal axial skeleton, abnormal somite development, abnormal neural tube morphology, abnormal developmental pattern) which makes sense, considering these phenotypes contain many of the same words, such as skeleton, somite, development, and morphology.

Level 2					
MGI ID	**Phenotype**	**TOP 30% % with term**	**Genome % with term**	**P-value**	**Adjusted P-value**
MP:0002006	tumorigenesis	98% (76/77)	9% (576/6325)	4.638E-76	1.391E-74
MP:0002873	normal phenotype	45% (35/77)	22% (1411/6325)	7.444E-06	0.0002233
MP:0003631	nervous system phenotype	57% (44/77)	37% (2377/6325)	0.0005621	0.017
MP:0005369	muscle phenotype	33% (26/77)	16% (1030/6325)	0.0001642	0.005
MP:0005370	liver/biliary system phenotype	36% (28/77)	13% (842/6325)	3.621E-07	0.00001086
MP:0005371	limbs/digits/tail phenotype	27% (21/77)	10% (678/6325)	0.00004951	0.001
MP:0005376	homeostasis/metabolism phenotype	67% (52/77)	42% (2719/6325)	0.00002216	0.0006648
MP:0005377	hearing/vestibular/ear phenotype	19% (15/77)	8% (516/6325)	0.001	0.042
MP:0005378	growth/size phenotype	74% (57/77)	41% (2608/6325)	8.134E-09	2.44E-07
MP:0005379	endocrine/exocrine gland phenotype	46% (36/77)	19% (1256/6325)	1.47E-07	4.409E-06

MP:0005380	embryogenesis phenotype	55% (43/77)	21% (1334/6325)	3.972E-11	1.192E-09
MP:0005381	digestive/alimentary phenotype	45% (35/77)	14% (939/6325)	2.124E-10	6.373E-09
MP:0005382	craniofacial phenotype	37% (29/77)	12% (797/6325)	2.872E-08	8.616E-07
MP:0005384	cellular phenotype	51% (40/77)	21% (1351/6325)	4.688E-09	1.406E-07
MP:0005385	cardiovascular system phenotype	55% (43/77)	25% (1639/6325)	5.372E-08	1.612E-06
MP:0005387	immune system phenotype	70% (54/77)	33% (2146/6325)	2.102E-10	6.307E-09
MP:0005388	respiratory system phenotype	35% (27/77)	15% (967/6325)	0.00001974	0.0005923
MP:0005389	reproductive system phenotype	50% (39/77)	22% (1421/6325)	1.219E-07	3.656E-06
MP:0005390	skeleton phenotype	46% (36/77)	17% (1121/6325)	6.01E-09	1.803E-07
MP:0005391	vision/eye phenotype	40% (31/77)	14% (922/6325)	4.621E-08	1.386E-06
MP:0005397	hematopoietic system phenotype	62% (48/77)	27% (1765/6325)	6.193E-10	1.858E-08
MP:0010768	mortality/aging	83% (64/77)	53% (3374/6325)	9.468E-08	0.00000284
MP:0010771	integument phenotype	48% (37/77)	19% (1223/6325)	1.62E-08	4.86E-07
Level 3					
MGI ID	**Phenotype**	**TOP 30% % with term**	**Genome % with term**	**P-value**	**Adjusted P-value**
MP:0002166	altered tumor susceptibility	93% (72/77)	8% (532/6325)	3.522E-69	2.817E-67
MP:0010639	altered tumor pathology	38% (30/77)	2% (147/6325)	1.024E-27	8.196E-26
MP:0002169	no abnormal phenotype detected	45% (35/77)	22% (1406/6325)	6.957E-06	0.0005566
MP:0003632	abnormal nervous system morphology	53% (41/77)	30% (1921/6325)	0.0000489	0.004
MP:0000516	abnormal urinary system morphology	24% (19/77)	10% (633/6325)	0.0001876	0.015
MP:0002108	abnormal muscle morphology	28% (22/77)	10% (658/6325)	9.468E-06	0.0007574
MP:0002138	abnormal hepatobiliary system morphology	35% (27/77)	11% (738/6325)	9.649E-08	0.00000772
MP:0002139	abnormal hepatobiliary system physiology	16% (13/77)	5% (329/6325)	0.000188	0.015
MP:0000545	abnormal limbs/digits/tail morphology	27% (21/77)	10% (678/6325)	0.00004951	0.004
MP:0005164	abnormal response to injury	20% (16/77)	5% (349/6325)	4.996E-06	0.0003997
MP:0008872	abnormal physiological response to xenobiotic	25% (20/77)	6% (421/6325)	1.407E-07	0.00001126
MP:0001270	distended abdomen	9% (7/77)	1% (71/6325)	0.00003682	0.003
MP:0002089	abnormal postnatal growth/weight/body size	57% (44/77)	32% (2047/6325)	0.00001205	0.0009636
MP:0004196	abnormal prenatal growth/weight/body size	45% (35/77)	13% (844/6325)	1.104E-11	8.835E-10
MP:0002163	abnormal gland morphology	46% (36/77)	17% (1131/6325)	7.634E-09	6.108E-07
MP:0001672	abnormal embryogenesis/ development	55% (43/77)	21% (1334/6325)	3.972E-11	3.177E-09
MP:0000462	abnormal digestive system morphology	37% (29/77)	11% (711/6325)	2.257E-09	1.805E-07
MP:0001663	abnormal digestive system physiology	22% (17/77)	7% (459/6325)	0.00003637	0.003
MP:0000428	abnormal craniofacial morphology	37% (29/77)	12% (797/6325)	2.872E-08	2.298E-06
MP:0005621	abnormal cell physiology	49% (38/77)	19% (1230/6325)	4.841E-09	3.873E-07

MP:0001544	abnormal cardiovascular system physiology	41% (32/77)	17% (1080/6325)	4.684E-07	0.00003747
MP:0002127	abnormal cardiovascular system morphology	51% (40/77)	20% (1278/6325)	8.808E-10	7.046E-08
MP:0000685	abnormal immune system morphology	54% (42/77)	24% (1518/6325)	1.114E-08	8.912E-07
MP:0001790	abnormal immune system physiology	61% (47/77)	26% (1698/6325)	4.609E-10	3.687E-08
MP:0002132	abnormal respiratory system morphology	25% (20/77)	9% (580/6325)	0.00001756	0.001
MP:0002160	abnormal reproductive system morphology	42% (33/77)	14% (945/6325)	5.011E-09	4.009E-07
MP:0005508	abnormal skeleton morphology	45% (35/77)	16% (1052/6325)	4.513E-09	3.61E-07
MP:0002092	abnormal eye morphology	37% (29/77)	13% (859/6325)	1.453E-07	0.00001162
MP:0002396	abnormal hematopoietic system morphology/development	62% (48/77)	26% (1706/6325)	1.379E-10	1.103E-08
MP:0003786	premature aging	7% (6/77)	0% (47/6325)	0.00003643	0.003
MP:0010769	abnormal survival	83% (64/77)	50% (3173/6325)	2.888E-09	2.311E-07
MP:0002060	abnormal skin morphology	35% (27/77)	10% (670/6325)	1.323E-08	1.058E-06
MP:0005501	abnormal skin physiology	18% (14/77)	3% (238/6325)	1.363E-06	0.000109
MP:0010678	abnormal skin adnexa morphology	35% (27/77)	9% (623/6325)	2.863E-09	2.29E-07
MP:0010680	abnormal skin adnexa physiology	14% (11/77)	2% (132/6325)	8.885E-07	0.00007108
Level 4					
MGI ID	**Phenotype**	**TOP 30% % with term**	**Genome % with term**	**P-value**	**adjusted P-value**
MP:0002019	abnormal tumor incidence		8% (523/6325)	5.67E-60	3.499E-57
MP:0010307	abnormal tumor latency		0% (31/6325)	2.887E-11	1.781E-08
MP:0000858	altered metastatic potential	23% (18/77)	1% (68/6325)	2.75E-18	1.697E-15
MP:0003448	altered tumor morphology	24% (19/77)	1% (105/6325)	1.39E-16	8.579E-14
MP:0002152	abnormal brain morphology	44% (34/77)	18% (1190/6325)	4.543E-07	0.0002803
MP:0003861	abnormal nervous system development	40% (31/77)	14% (914/6325)	3.78E-08	0.00002332
MP:0005620	abnormal muscle contractility	16% (13/77)	4% (275/6325)	0.00003235	0.02
MP:0010630	abnormal cardiac muscle tissue morphology	16% (13/77)	4% (257/6325)	0.0000163	0.01
MP:0000598	abnormal liver morphology	35% (27/77)	11% (725/6325)	6.725E-08	0.00004149
MP:0002109	abnormal limb morphology	23% (18/77)	6% (440/6325)	5.402E-06	0.003
MP:0002115	abnormal skeleton extremities morphology	19% (15/77)	5% (361/6325)	0.00003189	0.02
MP:0009115	abnormal fat cell morphology	10% (8/77)	1% (106/6325)	0.00005979	0.037
MP:0001784	abnormal fluid regulation	25% (20/77)	8% (521/6325)	3.672E-06	0.002
MP:0008873	increased physiological sensitivity to xenobiotic	16% (13/77)	2% (183/6325)	4.581E-07	0.0002827
MP:0001731	abnormal postnatal growth	29% (23/77)	11% (758/6325)	0.00002638	0.016
MP:0003956	abnormal body size	51% (40/77)	29% (1850/6325)	0.00004067	0.025

MP:0002088	abnormal embryonic growth/weight/body size	40% (31/77)	11% (711/6325)	9.161E-11	5.652E-08
MP:0004197	abnormal fetal growth/weight/body size	16% (13/77)	3% (208/6325)	1.802E-06	0.001
MP:0010865	prenatal growth retardation	31% (24/77)	7% (462/6325)	9.424E-10	5.815E-07
MP:0010866	abnormal prenatal body size	33% (26/77)	9% (618/6325)	1.156E-08	0.00000713
MP:0001697	abnormal embryo size	27% (21/77)	8% (506/6325)	5.678E-07	0.0003503
MP:0002084	abnormal developmental patterning	22% (17/77)	6% (431/6325)	0.00001657	0.01
MP:0002085	abnormal embryonic tissue morphology	33% (26/77)	11% (757/6325)	6.225E-07	0.0003841
MP:0002086	abnormal extraembryonic tissue morphology	36% (28/77)	9% (579/6325)	1.103E-10	6.806E-08
MP:0003886	abnormal embryonic epiblast morphology	7% (6/77)	0% (50/6325)	0.00005007	0.031
MP:0003984	embryonic growth retardation	29% (23/77)	6% (406/6325)	4.644E-10	2.865E-07
MP:0000477	abnormal intestine morphology	23% (18/77)	5% (321/6325)	6.367E-08	0.00003929
MP:0010352	gastrointestinal tract polyps	9% (7/77)	0% (21/6325)	2.669E-08	0.00001647
MP:0000432	abnormal head morphology	27% (21/77)	9% (590/6325)	6.275E-06	0.004
MP:0003935	abnormal craniofacial development	22% (17/77)	4% (294/6325)	1.042E-07	0.00006428
MP:0000313	abnormal cell death	35% (27/77)	8% (563/6325)	3.253E-10	2.007E-07
MP:0000350	abnormal cell proliferation	28% (22/77)	5% (331/6325)	6.728E-11	4.151E-08
MP:0000249	abnormal blood vessel physiology	15% (12/77)	3% (237/6325)	0.00003556	0.022
MP:0002128	abnormal blood circulation	25% (20/77)	8% (547/6325)	7.525E-06	0.005
MP:0002972	abnormal cardiac muscle contractility	14% (11/77)	2% (188/6325)	0.00002201	0.014
MP:0000266	abnormal heart morphology	41% (32/77)	13% (875/6325)	3.204E-09	1.977E-06
MP:0001614	abnormal blood vessel morphology	41% (32/77)	13% (848/6325)	1.48E-09	9.132E-07
MP:0002925	abnormal cardiovascular development	38% (30/77)	9% (629/6325)	2.456E-11	1.516E-08
MP:0000716	abnormal immune system cell morphology	48% (37/77)	19% (1247/6325)	2.748E-08	0.00001695
MP:0002722	abnormal immune system organ morphology	40% (31/77)	14% (927/6325)	5.233E-08	0.00003229
MP:0001819	abnormal immune cell physiology	37% (29/77)	17% (1102/6325)	0.00002792	0.017
MP:0001845	abnormal inflammatory response	41% (32/77)	12% (822/6325)	6.825E-10	4.211E-07
MP:0002420	abnormal adaptive immunity	37% (29/77)	17% (1104/6325)	0.00002859	0.018
MP:0002421	abnormal cell-mediated immunity	37% (29/77)	17% (1109/6325)	0.00003039	0.019
MP:0002723	abnormal immune serum protein physiology	31% (24/77)	12% (809/6325)	0.00002382	0.015
MP:0001175	abnormal lung morphology	22% (17/77)	7% (450/6325)	0.00002845	0.018
MP:0000653	abnormal sex gland morphology	28% (22/77)	9% (597/6325)	2.004E-06	0.001
MP:0001119	abnormal female reproductive system morphology	28% (22/77)	6% (436/6325)	9.68E-09	5.972E-06
MP:0001145	abnormal male reproductive system morphology	23% (18/77)	8% (529/6325)	0.00006092	0.038

MP:0000163	abnormal cartilage morphology	18% (14/77)	4% (299/6325)	0.00001708	0.011
MP:0002113	abnormal skeleton development	29% (23/77)	7% (484/6325)	1.21E-08	7.468E-06
MP:0002114	abnormal axial skeleton morphology	31% (24/77)	11% (711/6325)	2.702E-06	0.002
MP:0003795	abnormal bone structure	20% (16/77)	6% (400/6325)	0.00002612	0.016
MP:0009250	abnormal appendicular skeleton morphology	19% (15/77)	6% (384/6325)	0.0000635	0.039
MP:0005193	abnormal anterior eye segment morphology	16% (13/77)	4% (293/6325)	0.00006087	0.038
MP:0005195	abnormal posterior eye segment morphology	20% (16/77)	6% (431/6325)	0.00006282	0.039
MP:0001851	eye inflammation	7% (6/77)	0% (50/6325)	0.00005007	0.031
MP:0000689	abnormal spleen morphology	35% (27/77)	10% (684/6325)	2.033E-08	0.00001254
MP:0000703	abnormal thymus morphology	19% (15/77)	6% (392/6325)	0.00007968	0.049
MP:0002398	abnormal bone marrow cell morphology/ development	40% (31/77)	15% (972/6325)	1.541E-07	0.00009508
MP:0002429	abnormal blood cell morphology/ development	59% (46/77)	24% (1531/6325)	4.779E-11	2.949E-08
MP:0002080	prenatal lethality	61% (47/77)	26% (1673/6325)	2.886E-10	1.78E-07
MP:0002081	perinatal lethality	38% (30/77)	16% (1075/6325)	5.038E-06	0.003
MP:0002083	premature death	58% (45/77)	16% (1058/6325)	3.133E-16	1.933E-13
MP:0010770	preweaning lethality	71% (55/77)	43% (2750/6325)	1.201E-06	0.0007407
MP:0001191	abnormal skin condition	20% (16/77)	3% (242/6325)	4.574E-08	0.00002822
MP:0001216	abnormal epidermal layer morphology	22% (17/77)	4% (261/6325)	1.987E-08	0.00001226
MP:0003453	abnormal keratinocyte physiology	9% (7/77)	1% (65/6325)	0.00002175	0.013
MP:0000367	abnormal coat/ hair morphology	24% (19/77)	6% (402/6325)	3.365E-07	0.0002076
MP:0000377	abnormal hair follicle morphology	18% (14/77)	2% (154/6325)	8.534E-09	5.266E-06
MP:0000627	abnormal mammary gland morphology	12% (10/77)	2% (144/6325)	0.00001332	0.008
MP:0000427	abnormal hair cycle	9% (7/77)	0% (48/6325)	3.528E-06	0.002
Level 5					
MGI ID	**Phenotype**	**TOP 30% % with term**	**Genome % with term**	**P-value**	**Adjusted P-value**
MP:0002020	increased tumor incidence	79% (61/77)	7% (463/6325)	7.194E-53	1.423E-49
MP:0002052	decreased tumor incidence	20% (16/77)	1% (124/6325)	5.062E-12	1.001E-08
MP:0010308	decreased tumor latency	11% (9/77)	0% (24/6325)	9.976E-11	1.973E-07
MP:0001272	increased metastatic potential	18% (14/77)	0% (38/6325)	4.756E-16	9.408E-13
MP:0003447	decreased tumor growth/size	11% (9/77)	0% (57/6325)	6.967E-08	0.0001378
MP:0000913	abnormal brain development	27% (21/77)	9% (570/6325)	3.696E-06	0.007
MP:0000738	impaired muscle contractility	16% (13/77)	3% (215/6325)	2.557E-06	0.005
MP:0002972	abnormal cardiac muscle contractility	14% (11/77)	2% (188/6325)	0.00002201	0.044

MP:0010080	abnormal hepatocyte physiology	11% (9/77)	1% (95/6325)	3.584E-06	0.007
MP:0000547	short limbs	10% (8/77)	1% (92/6325)	0.00002314	0.046
MP:0000550	abnormal forelimb morphology	12% (10/77)	2% (139/6325)	9.946E-06	0.02
MP:0003723	abnormal long bone morphology	19% (15/77)	4% (315/6325)	6.689E-06	0.013
MP:0004499	increased incidence of chemically-induced tumors	11% (9/77)	1% (76/6325)	6.452E-07	0.001
MP:0001265	decreased body size	50% (39/77)	26% (1660/6325)	6.099E-06	0.012
MP:0003984	embryonic growth retardation	29% (23/77)	6% (406/6325)	4.644E-10	9.185E-07
MP:0001697	abnormal embryo size	27% (21/77)	8% (506/6325)	5.678E-07	0.001
MP:0001698	decreased embryo size	27% (21/77)	7% (497/6325)	4.257E-07	0.0008421
MP:0001674	abnormal triploblastic development	15% (12/77)	3% (207/6325)	9.786E-06	0.019
MP:0001688	abnormal somite development	15% (12/77)	3% (204/6325)	8.498E-06	0.017
MP:0001711	abnormal placenta morphology	20% (16/77)	5% (329/6325)	2.399E-06	0.005
MP:0001718	abnormal visceral yolk sac morphology	18% (14/77)	3% (195/6325)	1.38E-07	0.0002729
MP:0003229	abnormal vitelline vasculature morphology	20% (16/77)	2% (179/6325)	7.934E-10	1.569E-06
MP:0000488	abnormal intestinal epithelium morphology	14% (11/77)	2% (164/6325)	6.484E-06	0.013
MP:0000496	abnormal small intestine morphology	12% (10/77)	2% (131/6325)	6.077E-06	0.012
MP:0008011	intestine polyps	9% (7/77)	0% (18/6325)	1.115E-08	0.00002206
MP:0003743	abnormal facial morphology	23% (18/77)	7% (466/6325)	0.00001168	0.023
MP:0001648	abnormal apoptosis	20% (16/77)	5% (374/6325)	0.00001165	0.023
MP:0008942	abnormal induced cell death	14% (11/77)	2% (163/6325)	6.135E-06	0.012
MP:0000352	decreased cell proliferation	20% (16/77)	3% (251/6325)	7.402E-08	0.0001464
MP:0001914	hemorrhage	23% (18/77)	7% (444/6325)	6.106E-06	0.012
MP:0005140	decreased cardiac muscle contractility	14% (11/77)	2% (154/6325)	3.663E-06	0.007
MP:0000288	abnormal pericardium morphology	19% (15/77)	3% (243/6325)	2.995E-07	0.0005925
MP:0010545	abnormal heart layer morphology	20% (16/77)	5% (344/6325)	4.179E-06	0.008
MP:0000259	abnormal vascular development	31% (24/77)	7% (451/6325)	5.895E-10	1.166E-06
MP:0000267	abnormal heart development	20% (16/77)	4% (297/6325)	6.57E-07	0.001
MP:0005460	abnormal leukopoiesis	32% (25/77)	12% (768/6325)	2.949E-06	0.006
MP:0000689	abnormal spleen morphology	35% (27/77)	10% (684/6325)	2.033E-08	0.00004021
MP:0002221	abnormal lymph organ size	31% (24/77)	10% (687/6325)	1.486E-06	0.003
MP:0001846	increased inflammatory response	37% (29/77)	10% (663/6325)	4.577E-10	9.052E-07
MP:0002442	abnormal leukocyte physiology	37% (29/77)	17% (1083/6325)	0.00002294	0.045
MP:0000627	abnormal mammary gland morphology	12% (10/77)	2% (144/6325)	0.00001332	0.026
MP:0009208	abnormal female genitalia morphology	19% (15/77)	5% (332/6325)	0.0000123	0.024
MP:0000166	abnormal chondrocyte morphology	11% (9/77)	1% (78/6325)	7.884E-07	0.002

MP:0003055	abnormal long bone epiphyseal plate morphology	14% (11/77)	2% (144/6325)	1.982E-06	0.004
MP:0000164	abnormal cartilage development	14% (11/77)	2% (152/6325)	3.252E-06	0.006
MP:0006395	abnormal epiphyseal plate morphology	16% (13/77)	2% (157/6325)	8.649E-08	0.0001711
MP:0008271	abnormal bone ossification	16% (13/77)	3% (240/6325)	0.00000807	0.016
MP:0002224	abnormal spleen size	27% (21/77)	8% (511/6325)	6.644E-07	0.001
MP:0002414	abnormal myeloblast morphology/ development	35% (27/77)	12% (770/6325)	2.267E-07	0.0004484
MP:0002123	abnormal hematopoiesis	59% (46/77)	23% (1515/6325)	3.28E-11	6.487E-08
MP:0008246	abnormal leukocyte morphology	48% (37/77)	19% (1237/6325)	2.209E-08	0.00004369
MP:0006208	lethality throughout fetal growth and development	22% (17/77)	6% (393/6325)	5.089E-06	0.01
MP:0008762	embryonic lethality	55% (43/77)	19% (1243/6325)	3.616E-12	7.153E-09
MP:0002080	prenatal lethality	61% (47/77)	26% (1673/6325)	2.886E-10	5.708E-07
MP:0010832	lethality during fetal growth through weaning	55% (43/77)	30% (1898/6325)	3.983E-06	0.008
MP:0010300	increased skin tumor incidence	14% (11/77)	0% (58/6325)	3.812E-10	7.541E-07
MP:0001222	epidermal hyperplasia	10% (8/77)	1% (72/6325)	4.372E-06	0.009
MP:0001212	skin lesions	10% (8/77)	1% (91/6325)	0.0000215	0.043
MP:0009582	abnormal keratinocyte proliferation	7% (6/77)	0% (37/6325)	0.00001064	0.021
MP:0001510	abnormal coat appearance	20% (16/77)	6% (393/6325)	0.00002116	0.042
MP:0000379	decreased hair follicle number	9% (7/77)	0% (46/6325)	2.732E-06	0.005
MP:0003704	abnormal hair follicle development	9% (7/77)	0% (56/6325)	8.907E-06	0.018

Table 2: Enriched Mammalian Phenotypes Identified in the Most Conserved (Top 30%) Tumorigenesis Orthologs.

Level 2					
MGI ID	**Phenotype**	**BOTTOM 30% with term**	**Genome % with term**	**P-value**	**Adjusted P-value**
MP:0002006	tumorigenesis	97% (75/77)	9% (577/6325)	1.76E-73	5.28E-72
MP:0005376	homeostasis/metabolism phenotype	63% (49/77)	43% (2722/6325)	0.00044	0.013
MP:0005379	endocrine/exocrine gland phenotype	41% (32/77)	19% (1260/6325)	1.93E-05	0.000578
MP:0005381	digestive/alimentary phenotype	31% (24/77)	15% (950/6325)	0.00034	0.01
MP:0005384	cellular phenotype	40% (31/77)	21% (1360/6325)	0.000229	0.007
MP:0005387	immune system phenotype	68% (53/77)	33% (2147/6325)	6.67E-10	2E-08
MP:0005389	reproductive system phenotype	40% (31/77)	22% (1429/6325)	0.000543	0.016
MP:0005397	hematopoietic system phenotype	54% (42/77)	28% (1771/6325)	1.3E-06	3.9E-05
MP:0010771	integument phenotype	40% (31/77)	19% (1229/6325)	0.000032	0.00096
Level 3					
MGI ID	**Phenotype**	**BOTTOM 30% % with term**	**Genome % with term**	**P-value**	**adjusted P-value**

MP:0002166	altered tumor susceptibility	89% (69/77)	8% (535/6325)	4.81E-63	3.85E-61
MP:0010639	altered tumor pathology	29% (23/77)	2% (154/6325)	2.3E-18	1.84E-16
MP:0002139	abnormal hepatobiliary system physiology	19% (15/77)	5% (327/6325)	1.03E-05	0.000826
MP:0005164	abnormal response to injury	19% (15/77)	5% (350/6325)	2.25E-05	0.002
MP:0008872	abnormal physiological response to xenobiotic	32% (25/77)	6% (416/6325)	1.87E-11	1.5E-09
MP:0002163	abnormal gland morphology	37% (29/77)	17% (1138/6325)	7.02E-05	0.006
MP:0002164	abnormal gland physiology	19% (15/77)	6% (422/6325)	0.000177	0.014
MP:0001663	abnormal digestive system physiology	22% (17/77)	7% (459/6325)	3.64E-05	0.003
MP:0005621	abnormal cell physiology	40% (31/77)	19% (1237/6325)	3.45E-05	0.003
MP:0000685	abnormal immune system morphology	53% (41/77)	24% (1519/6325)	5.07E-08	4.06E-06
MP:0001790	abnormal immune system physiology	58% (45/77)	26% (1700/6325)	7.96E-09	6.37E-07
MP:0002160	abnormal reproductive system morphology	33% (26/77)	15% (952/6325)	4.26E-05	0.003
MP:0002396	abnormal hematopoietic system morphology/ development	54% (42/77)	27% (1712/6325)	4.72E-07	3.78E-05
MP:0001657	abnormal induced morbidity/mortality	22% (17/77)	6% (409/6325)	8.51E-06	0.000681
MP:0002060	abnormal skin morphology	27% (21/77)	10% (676/6325)	4.75E-05	0.004
MP:0005501	abnormal skin physiology	15% (12/77)	3% (240/6325)	4E-05	0.003
MP:0010678	abnormal skin adnexa morphology	27% (21/77)	9% (629/6325)	1.65E-05	0.001
Level 4					
MGI ID	**Phenotype**	**BOTTOM 30% % with term**	**Genome % with term**	**P-value**	**adjusted P-value**
MP:0002019	abnormal tumor incidence	87% (67/77)	8% (523/6325)	5.67E-60	3.5E-57
MP:0010307	abnormal tumor latency	9% (7/77)	0% (34/6325)	4.47E-07	0.000276
MP:0000858	altered metastatic potential	14% (11/77)	1% (75/6325)	4.35E-09	2.68E-06
MP:0003448	altered tumor morphology	23% (18/77)	1% (106/6325)	2.64E-15	1.63E-12
MP:0000609	abnormal liver physiology	18% (14/77)	4% (299/6325)	1.71E-05	0.011
MP:0005023	abnormal wound healing	11% (9/77)	1% (83/6325)	1.27E-06	0.000785
MP:0008873	increased physiological sensitivity to xenobiotic	19% (15/77)	2% (181/6325)	7.58E-09	4.67E-06
MP:0008874	decreased physiological sensitivity to xenobiotic	18% (14/77)	2% (184/6325)	7E-08	4.32E-05
MP:0000477	abnormal intestine morphology	18% (14/77)	5% (325/6325)	4.17E-05	0.026
MP:0010155	abnormal intestine physiology	12% (10/77)	2% (177/6325)	7.12E-05	0.044
MP:0000313	abnormal cell death	24% (19/77)	9% (571/6325)	4.87E-05	0.03
MP:0010094	abnormal chromosome stability	9% (7/77)	1% (74/6325)	4.71E-05	0.029
MP:0002925	abnormal cardiovascular development	27% (21/77)	10% (638/6325)	2.03E-05	0.013
MP:0000716	abnormal immune system cell morphology	48% (37/77)	19% (1247/6325)	2.75E-08	1.7E-05
MP:0002722	abnormal immune system organ morphology	40% (31/77)	14% (927/6325)	5.23E-08	3.23E-05

MP:0001819	abnormal immune cell physiology	44% (34/77)	17% (1097/6325)	5.22E-08	3.22E-05
MP:0001845	abnormal inflammatory response	32% (25/77)	13% (829/6325)	1.13E-05	0.007
MP:0002419	abnormal innate immunity	23% (18/77)	6% (438/6325)	5.08E-06	0.003
MP:0002420	abnormal adaptive immunity	44% (34/77)	17% (1099/6325)	5.46E-08	3.37E-05
MP:0002421	abnormal cell-mediated immunity	45% (35/77)	17% (1103/6325)	1.56E-08	9.63E-06
MP:0002452	abnormal antigen presenting cell physiology	35% (27/77)	11% (732/6325)	8.18E-08	5.05E-05
MP:0005025	abnormal response to infection	27% (21/77)	6% (441/6325)	6.04E-08	3.73E-05
MP:0002405	respiratory system inflammation	12% (10/77)	2% (180/6325)	8.15E-05	0.05
MP:0001119	abnormal female reproductive system morphology	22% (17/77)	6% (441/6325)	2.21E-05	0.014
MP:0000689	abnormal spleen morphology	31% (24/77)	10% (687/6325)	1.49E-06	0.000917
MP:0002398	abnormal bone marrow cell morphology/ development	36% (28/77)	15% (975/6325)	8.48E-06	0.005
MP:0002429	abnormal blood cell morphology/ development	49% (38/77)	24% (1539/6325)	3.01E-06	0.002
MP:0002083	premature death	36% (28/77)	16% (1075/6325)	4.88E-05	0.03
MP:0001191	abnormal skin condition	19% (15/77)	3% (243/6325)	3E-07	0.000185
MP:0000627	abnormal mammary gland morphology	16% (13/77)	2% (141/6325)	2.65E-08	1.63E-05
Level 5					
MGI ID	**Phenotype**	**BOTTOM 30% % with term**	**Genome % with term**	**P-value**	**adjusted P-value**
MP:0002020	increased tumor incidence	76% (59/77)	7% (465/6325)	1.38E-49	2.72E-46
MP:0002052	decreased tumor incidence	22% (17/77)	1% (123/6325)	3.61E-13	7.14E-10
MP:0010308	decreased tumor latency	7% (6/77)	0% (27/6325)	2.13E-06	0.004
MP:0001273	decreased metastatic potential	7% (6/77)	0% (32/6325)	5.05E-06	0.01
MP:0003447	decreased tumor growth/size	12% (10/77)	0% (56/6325)	4.27E-09	8.45E-06
MP:0003721	increased tumor growth/size	7% (6/77)	0% (40/6325)	1.59E-05	0.031
MP:0002908	delayed wound healing	9% (7/77)	0% (39/6325)	1.02E-06	0.002
MP:0004499	increased incidence of chemically-induced tumors	16% (13/77)	1% (72/6325)	1.44E-11	2.84E-08
MP:0004502	decreased incidence of chemically-induced tumors	11% (9/77)	0% (42/6325)	6.62E-09	1.31E-05
MP:0005460	abnormal leukopoiesis	32% (25/77)	12% (768/6325)	2.95E-06	0.006
MP:0000689	abnormal spleen morphology	31% (24/77)	10% (687/6325)	1.49E-06	0.003
MP:0002221	abnormal lymph organ size	36% (28/77)	10% (683/6325)	4.34E-09	8.58E-06
MP:0001846	increased inflammatory response	29% (23/77)	10% (669/6325)	3.48E-06	0.007
MP:0002442	abnormal leukocyte physiology	44% (34/77)	17% (1078/6325)	3.37E-08	6.67E-05
MP:0002452	abnormal antigen presenting cell physiology	35% (27/77)	11% (732/6325)	8.18E-08	0.000162
MP:0002459	abnormal B cell physiology	24% (19/77)	7% (482/6325)	4.71E-06	0.009

MP:0001793	altered susceptibility to infection	25% (20/77)	6% (409/6325)	8.92E-08	0.000176
MP:0001861	lung inflammation	12% (10/77)	2% (154/6325)	2.31E-05	0.046
MP:0000627	abnormal mammary gland morphology	16% (13/77)	2% (141/6325)	2.65E-08	5.24E-05
MP:0002224	abnormal spleen size	27% (21/77)	8% (511/6325)	6.64E-07	0.001
MP:0002123	abnormal hematopoiesis	49% (38/77)	24% (1523/6325)	2.74E-06	0.005
MP:0008246	abnormal leukocyte morphology	46% (36/77)	19% (1238/6325)	1.17E-07	0.000232
MP:0000630	mammary gland hyperplasia	6% (5/77)	0% (11/6325)	8.7E-07	0.002

Table 3: Enriched Mammalian Phenotypes Identified in the Most Divergent (Bottom 30%) Tumorigenesis Orthologs.

However, a less intuitive connection is represented by the magenta highlighted phenotypes that overlap with the light-blue highlighted phenotypes at the bottom left corner (abnormal neural tube morphology, abnormal developmental pattern, cleft palate, abnormal palate morphology, abnormal cranial morphology) which form a connection with the set of magenta highlighted phenotypes on the right lower edge of the heat map (abnormal purkinje cell morphology, abnormal sensory neuron, abnormal brain ventricle). Together, these light-blue and magenta highlighted enriched phenotypes elucidate an important developmental relationship encoded by these tumorigenesis orthologs that controls axial skeletal morphology, cranial morphology, facial morphology, facial development, cleft palate, neural tube morphology, somite development, purkinje cell morphology, sensory neurons, and brain ventricles. It is not surprising that somite development, neural tube development, cranial morphology, and facial morphology represents inter-related phenotypes that must be very tightly regulated in order to produce functional heads, brains and faces in during embryogenesis.

Annotation term co-occurrence in PubMed and visualization of resulting network

The same enriched gene ontology and mammalian phenotype annotation terms that produced the co-occurrence heat maps were also used to produce networks representing the co-occurrence relationships between pairs of enriched terms in PubMed abstracts. The networks are composed of nodes (circles) and edges (arrows connecting nodes). Each node represents a specific enriched term and the arrow between two nodes points to the second (co-occurring) term. The number along the length of the arrow shows the number of abstracts produced by two different queries: (query-1) termX AND termY (smaller of the two numbers); (query-2) termX OR termY (larger of the two terms).

The enriched gene ontology terms were used to produce the network in Figure 7A. This network provides insight into the relationships that exist between the signaling pathways and embryological development. For example, the node at the top of this network is 'organ morphogenesis' which is connected to nodes associated with odontogenesis, uroteric bud, epithelial morphogenesis, branching morphogenesis, and cell fate commitment. Other portions of this network include gene ontology terms related to skeletal development including osteoclast, osteoclast development, bone morphogenetic protein, BMP signaling, osteoblast differentiation and ossification. Within this portion of the network are three cancer nodes: osteosarcoma, pancreatic cancer, and hepatocellular carcinoma.

Within the upper left region of the network are nodes associated with mononuclear cells including leukocyte proliferation, T cell proliferation, lymphocyte proliferation and mononuclear cell proliferation. Directly connected to these nodes are four cancer nodes (myeloma, multiple myeloma, myeloid leukemia, and chronic myeloid leukemia). Immediately connected to these nodes are a node associated with macrophage activation and a node associated with cell cycle progression. T-cell receptor and T cell receptor signaling connected to macrophage activation and cell cycle progression. And within this region is a third cluster of cancer nodes including medullablastoma, glioma, bladder cancer, small cell lung carcinoma, non small cell lung carcinoma, thyroid cancer, endometrial cancer and basal cell carcinoma. Within this network region are multiple nodes associated with specific signaling pathways including, mTOR, p53 signaling, jak- stat, WNT signaling, cytokine-cytokine interactions, TGF beta, and adherens junctions.

Figure 7B contains a network derived from mammalian phenotype terms. This network contains many distributed nodes associated with fetal growth and development including oogenesis, growth retardation, growth arrest, embryonic lethality, and abnormal development. Nodes to connected to these embryological nodes include cardiovascular system development, brain development, facial morphology, abnormal facial development, and somite development. A cluster of nodes is associated with bone ossification such as epiphyseal plate, chondrocyte, abnormal axial skeleton, short limbs and abnormal bone ossification. The strong embryogenesis theme is centered on head, brain, face and skull development as well as skeletal and cardiovascular developmental programs.

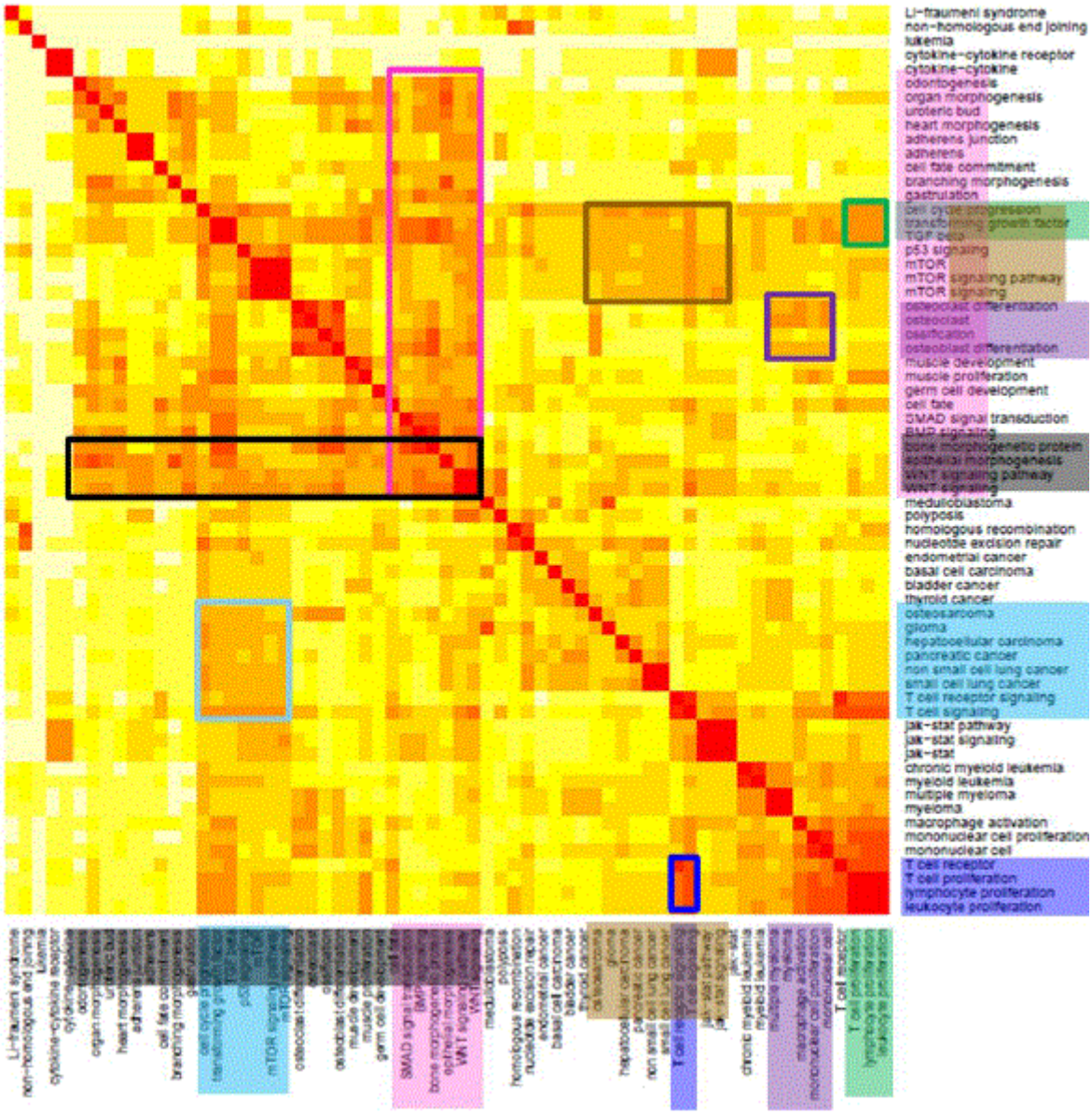

Figure 6A: PubMed Co-Occurrence Heat Map for Gene Ontology Enriched in 256 Tumorigenesis Orthologs. A representative set of enriched gene ontology terms were selected to query PubMed via the literature blasting interface, PubAtlas. The selected terms were chosen to reflect the identified themes (for example embryogenesis, tumorigenesis, anatomical morphology, organogenesis, immune system). The heat map displays the strongest associations between co-occurring terms in red, while weaker associations are represented by a continuum of color ranging from orange (strong associations), to yellow (moderate associations) and ultimately white (no associations). To aide in the visualization of connections between the terms, rectangular outlines were placed on the heat map and the corresponding terms were colored with the same color as the rectangle outlining the heat map colored pixels. Relationships between tumorigenesis processes and skeleton development and morphology (the regulation of osteoclast and osteoblast differentiation) as well as immune system function (proliferation of lymphocytes, leukocytes and T-cells) were identified.

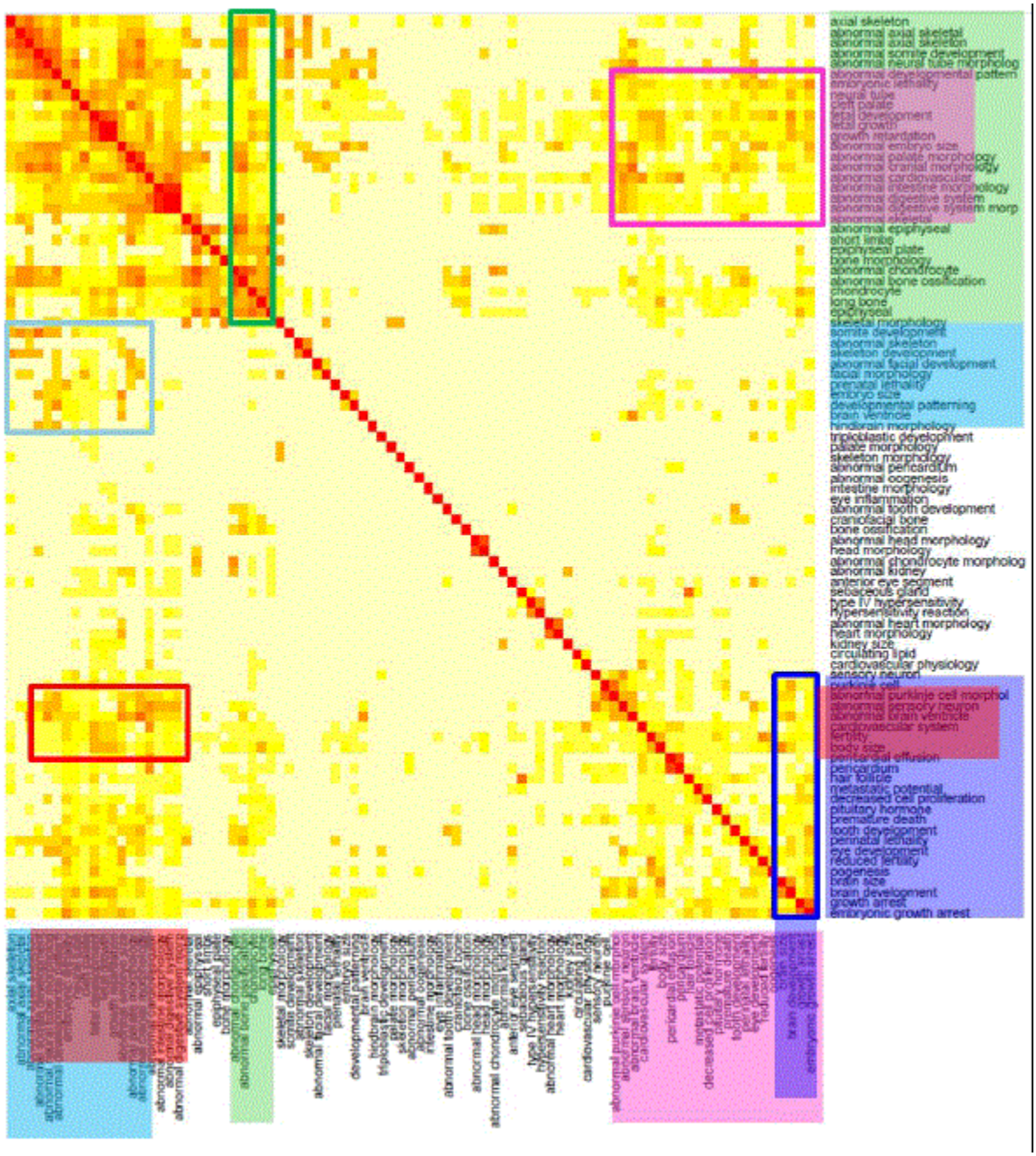

Figure 6B: PubMed Co-Occurrence Heat Map for Phenotypes Enriched in 256 Tumorigenesis Orthologs. Connections between enriched mammalian phenotype terms are visualized in this heat map. Themes identified include skeletal morphology, somite development, abnormal skeleton, skeleton development, abnormal facial development, facial morphology as well as axial skeleton, abnormal axial skeleton, abnormal somite development, abnormal neural tube morphology, abnormal developmental pattern are detected. The phenotype themes parallel many of the gene ontology themes that were also detected.

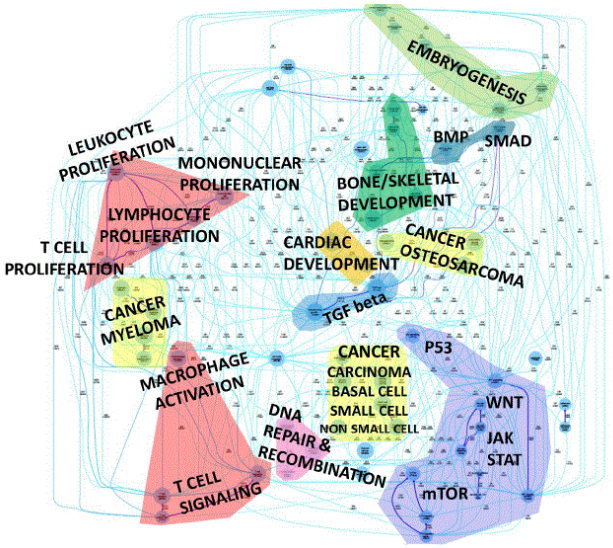

Figure 7A: Co-Occurrence Network for 'Gene Ontology Annotation Terms' in the Tumorigenesis Orthologs. This network, produced from enriched gene ontology terms, provides insight into the relationships that exist between the signaling pathways and embryological development. For example, the node at the top of this network is 'organ morphogenesis' which is connected to nodes associated with odontogenesis, uroteric bud, epithelial morphogenesis, branching morphogenesis, and cell fate commitment. Other portions of this network include gene ontology terms related to skeletal development including osteoclast, osteoclast development, bone morphogenetic protein, BMP signaling, osteoblast differentiation and ossification. Within this portion of the network are three cancer nodes: osteosarcoma, pancreatic cancer, and hepatocellular carcinoma.

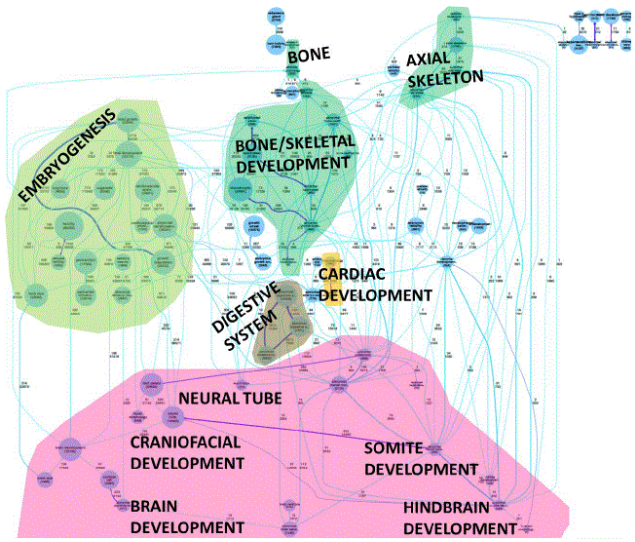

Figure 7B: Co-Occurrence Network for 'Mammalian Phenotype Annotation Terms' in the Tumorigenesis Orthologs. The visualized network is derived from mammalian phenotype terms. This network contains many distributed nodes associated with fetal growth and development including oogenesis, growth retardation, growth arrest, embryonic lethality, and abnormal development. Nodes to connected to these embryological nodes include cardiovascular system development, brain development, facial morphology, abnormal facial development, and somite development. A cluster of nodes is associated with bone ossification such as epiphyseal plate, chondrocyte, abnormal axial skeleton, short limbs and abnormal bone ossification. The strong embryogenesis theme is centered on head, brain, face and skull development as well as skeletal and cardiovascular developmental programs.

Biocarta pathways enriched for conserved and divergent tumorigenesis orthologs

To further identify specific cellular signaling pathways that regulate the phenotypes and biological processes represented in the PubMed co-occurrence networks, the tumorigenesis orthologs were analyzed to identify enrich pathways. The first analysis performed was an enrichment analysis on the complete set of 256 orthologs. This analysis identified the cell cycle pathway regulated by TGF beta as highly enriched for tumorigenesis orthologs (Figure 8A) as well as the cyclin and cell cycle regulation pathway (Figure 8B). Next enrichment analysis was performed on TOP 30% most conserved tumorigenesis orthologs which identified the TGF beta signaling pathway (Figure 9A) and the pathway associated with NFKB activation by nontypeable hemophilus influenza (Figure 9B). Finally, the pathway enrichment was applied to the BOTTOM 30% least conserved orthologs resulting in the identification of the cytokine network pathway (Figure 10A) and the cytokine and inflammation response pathway (Figure 10B). These pathways provide further evidence for the roles of the conserved and divergent orthologs.

Visualization of tumorigenesis ortholog knowledge extracted from pubmed abstracts

Queries were generated to identify pubmed abstracts associated with specific subsets of the tumorigenesis orthologs and focused on a particular area of biological knowledge. Figure 11 shows knowledge-mined visualizations for abstracts associated with genetic variation search terms coupled with either the TOP 30% most conserved orthologs, or the BOTTOM 30% least conserved orthologs. Within the TOP 30% most conserved orthologs, a number of results related to canine musculo-skeletal disorders, including osteoarthritis, hip dysplasia, dysplastic, and hip laxity. Genes that were identified included COL1A1, COL1A2, SLC3A1, SLC7A9, and RUNX2. The same pubmed query was used to repeat the query with the BOTTOM 30% least conserved orthologs. In contrast to the musculoskeletal rich results obtained with the TOP 30% orthologs, the results obtained for the least conserved orthologs showed associations with BRCA1, BRCA2, CYP1A2, cyclinD1, Interleukin 6 and the following disorders: tumor, breast cancer, mammary tumors, osteosarcoma, hemangiomas, mast cell tumors and cancer.

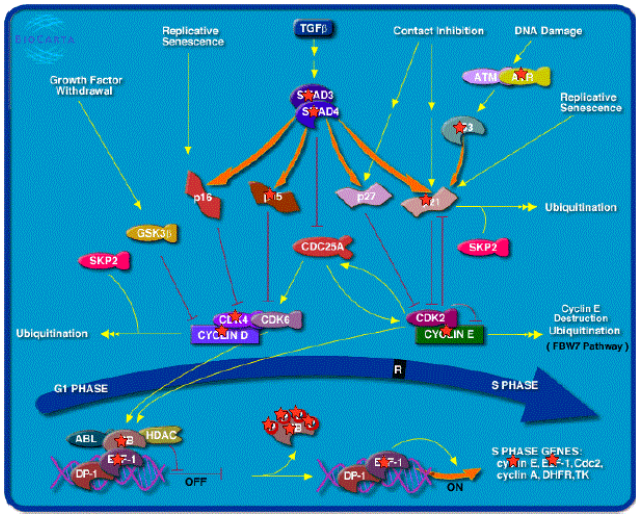

Figure 8A: Cell Cycle Pathway Enriched for Genes from the Complete Set of 256 Tumorigenesis Orthologs. The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the complete set of 256 orthologs and identified the cell cycle pathway regulated by TGF beta as highly enriched for tumorigenesis orthologs.

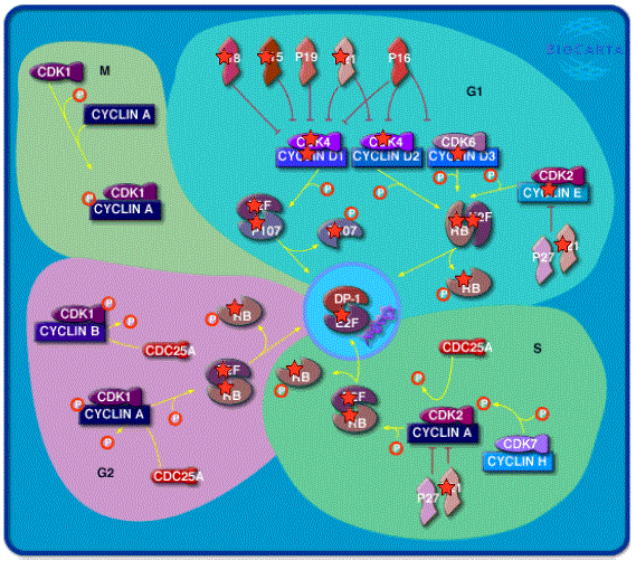

Figure 8B: Cyclins and Cell Cycle Regulation Pathway Enriched for Genes from the Complete Set of 256 Tumorigenesis Orthologs. The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the complete set of 256 orthologs and identified the the cyclin and cell cycle regulation pathway as highly enriched for tumorigenesis orthologs.

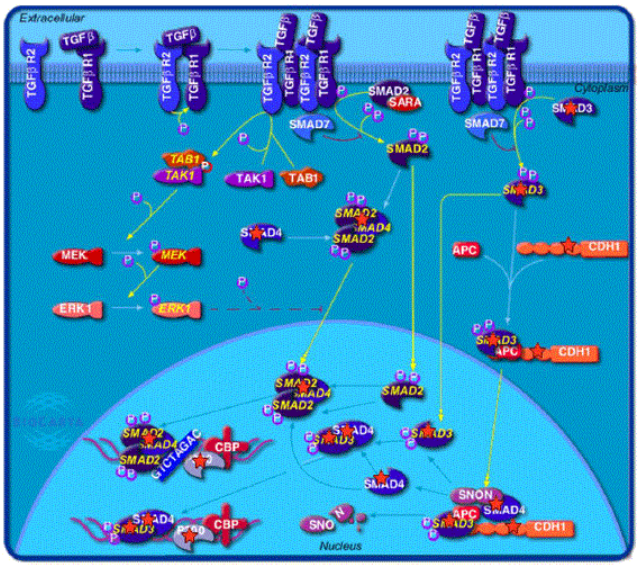

Figure 9A: GF beta Signaling Pathway is Enriched for Genes Identified in the Most Conserved Orthologs (30% Top). The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the TOP 30% most conserved orthologs and identified the TGF beta signaling pathway as being enriched for most conserved orthologs.

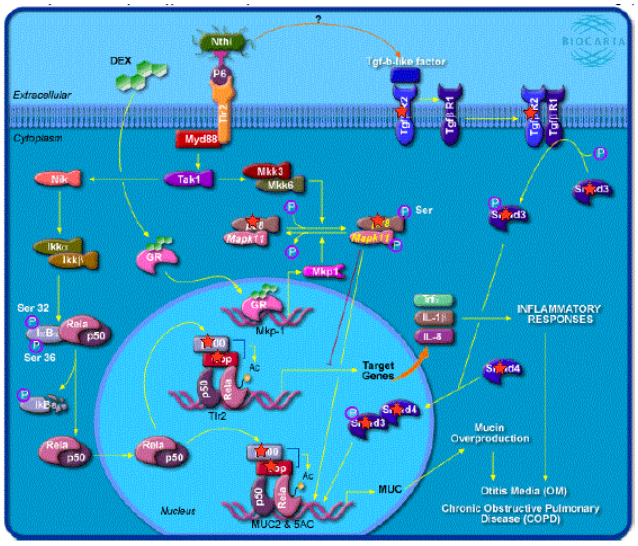

Figure 9B: Identified in the Most Conserved Orthologs (Top 30%). The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the TOP 30% most conserved orthologs and identified the NFKB Pathway Activation by Nontypeable Hemophilus Influenzae as being enriched for most conserved orthologs.

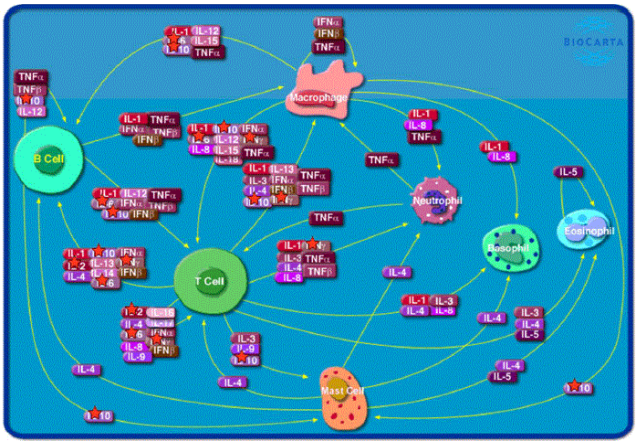

Figure 10A: Cytokine Network Pathway is Enriched for Genes Identified in Most Divergent Orthologs (Bottom 30%). The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the BOTTOM 30% least conserved orthologs and identified the cytokine network pathway as being enriched for least conserved orthologs.

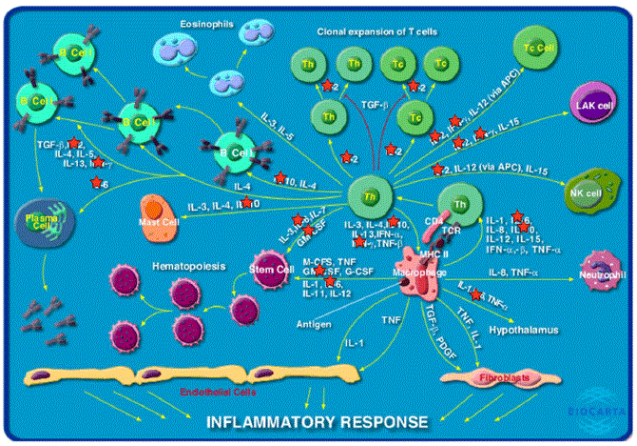

Figure 10B: Cytokines and Inflammatory Response Pathway is Enriched for Genes Identified in Most Divergent Orthologs (Bottom 30%). The tumorigenesis orthologs were analyzed to identify enriched pathways. This analysis was performed on the BOTTOM 30% least conserved orthologs and identified the cytokine and inflammation response pathway as being enriched for least conserved orthologs.

The pubmed data mining analysis was repeated on the same sets of orthologs, however, the query was changed to a focus on cellular differentiation and cell migration as well as hematopoeisis. A total of 200 pubmed abstracts from review papers were mined to identify the most over-represented terms across all of the abstracts. Results obtained for the TOP 30% most conserved orthologs included hematopoiesis, Smad2, Fox01 Wnt4, mTOR, TGFbeta, PDGFRa, Wnt7b and Runx1, Smad4, Dlx3 and Beta catenin. The analysis in the BOTTOM 30% only produced 61 abstracts. After mining the abstracts, the following overrepresented terms were identified MMP9, MMP2, BMPR1a, BMP4, Cyp2E1, Cyp1A1, Cyp2B1 along with Brca1, cyclin D1, and cyclin A.

The next set of queries was focused on terms relating to the tumor micro-environment applied to both the TOP 30% and BOOTOM 30% orthologs (Figure 12). Within the TOP 30% orthologs, the following terms were identified: inflammation, HIF1a, IL6, Bcl2, TGFbeta, TGFbeta receptor 2, Myeloma, Metastasis, Ecadherin, colon cancer, IL8, IL10, VEGF, hypoxia, hepatoma, neuroblastoma, melanoma and CXCR4. When query was repeated with the BOTTOM 30% of orthologs, the following terms were identified: lung cancer, inflammation, hypoxia, stat3, IL8, FoxP3, CCL5, CCR2, IL17A, CXCL9, TNFalpha, IFNgamma, breast cancer, CXCL7, NOS, IL22 and TP53.

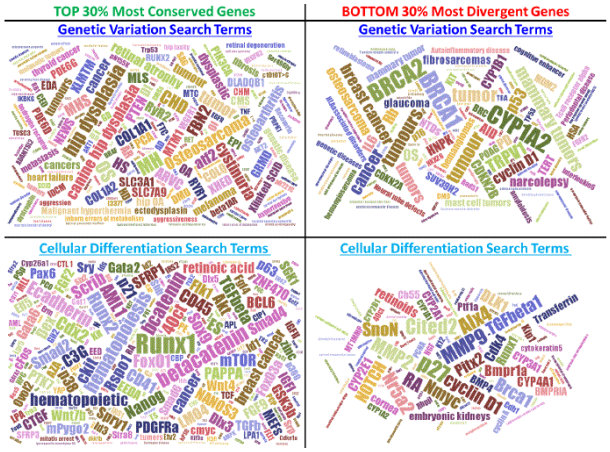

Figure 11: Word Cloud Visualization of 'Genes-Disease-Polymorphisms' PubMed Queries of Conserved and Divergent Tumorigenesis Orthologs Visualizations for abstracts associated with genetic variation search terms coupled with either the TOP 30% most conserved orthologs (shown on the upper left), or the BOTTOM 30% least conserved orthologs (shown on the upper right). Within the TOP 30% most conserved orthologs, a number of results related to canine musculo-skeletal disorders, including osteoarthritis, hip dysplasia, dysplastic, and hip laxity. Visualizations for abstracts associated with cellular differentiation and cell migration as well as hematopoiesis are shown on the bottom left and bottom right for TOP 30% and BOTTOM 30% (respectively).

Next a query designed to retrieve abstracts relating to reactive oxyegen species and reactive nitrogen species was executed across all 256 orthologs, split into TOP 30%, MIDDLE 40%, and BOTTOM 30% (Figure 12). Themes identified in the TOP 30% included cancer, hypoxia, toxicity, metastasis, HIF1a, melatonin, COX2, ATM, miR21, p21 as well as p53. Within the MIDDLE 40% orthologs, the following terms were identified carcinogenesis, inflammation, lung cancer, breast cancer, malignancies, iNOS, ROS, MMP9, TGFbeta and oxygen. Within the BOTTOM 30% of orthologs, the query and subsdequent knowledge mining produced a set of 200 abstracts. Over represented terms include tumor, BRCA1, resveratrol, melatonin, arsenic, estrogens, metastasis, gastric cancer, p53, Bcl2, lung toxicity, flavonoids, selenium, and oxygen.

The next query mining focus was myeloid lineage. Within the set of TOP 30% conserved orthologs, identified terms included

inflammation, hypoxia, miR21, IL18, osteoporosis, atopic dermatitis, IFNgamma, VEGF, asthma, atherosclerosis, cancer, TLR4, EGFR, CRH, IL1b, COX2, melanoma, RUNX1, and insulin. The results obtained from the BOTTOM 30% (least conserved orthologs) incude obesity, insulin, infection, vasculitis, polyangiitis, metastasis and the following genes: IL10, IL17, FOxm1, Cxcr4, Cxcl12, cscl10, IL1Ra, CxCL4, GDF15, CXCL1, IL17a, CXCL8. The last queries were focused on lymphoid lineage (Figure 13). The most represented term s associated with the top 30% most conserved orthologs are: inflammation, autoimmune disease, loymphomas, multiple sclerosis, inflammatory disease and psoriasis. The notable genes were Lef1, mTOR, IL22, Bcl2, IL23, MYD88, IL4, IL6, IRF8, and caspase 1.

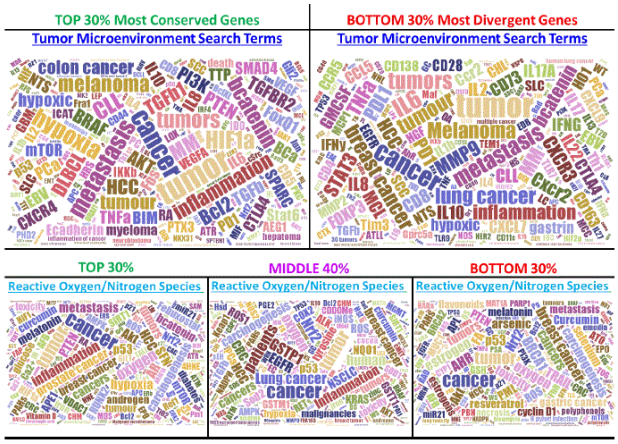

Figure 12: Word Cloud View of 'Tumor-Environment' and 'Reactive Oxygen/Nitrogen Species' Queries of Tumorigenesis Orthologs. A set of queries was focused on terms relating to the tumor micro-environment applied to both the TOP 30% (top left side) and BOOTOM 30% orthologs (top right side). Underneath, visualizations for a query designed to retrieve abstracts relating to reactive oxyegen species and reactive nitrogen species was executed across all 256 orthologs, split into TOP 30% (bottom left), MIDDLE 40% (bottom middle), and BOTTOM 30% (bottom right).

Within the BOTTOM 30% least conserved genes, the following disorders were identified: multiple sclerosis, arthritis, depression, infection, lymphoma, carcinoma, and infection. The following genes were identified: IL15, CXCL10, IL22, IL21, FoxP3, Cyclin D1, CXCR2, and CD8.

Characterization of single nucleotide polymorphisms in tumorigenesis orthologs

The number of SNPs deposited in dbSNP for tumorigenesis orthologs was compared to the average protein pair-wise identity. Figure 14A shows the correlation for dog SNPs. For tumorigenesis orthologs having high protein percent identity, approximately 5 SNPs were present in the dbSNP database. As the pairwise percent identity decreased, the number of SNPs increased above 20 for orthologs with very low percent identity. Figure 14B shows the scatter plot and correlation for SNPs in mouse genes. Similar to the pattern observed in dogs, the mouse orthologs having the highest percent identity contain the fewest SNPS. For example, mouse orthologs having the highest percent identity have approximately 100 SNPs while the mouse orthologs approaching 60% identity have approximately 400 SNPs. Figure 14C shows the scatter plot for human tumorigenesis orthologs. A similar inverse relationship between ortholog pair-wise percent identity and number of SNPs is observed.

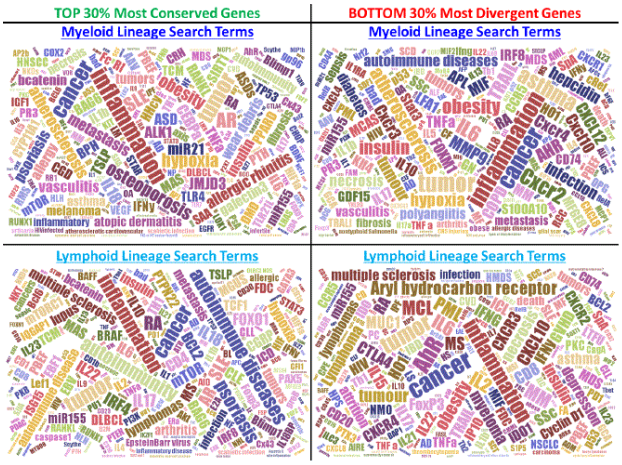

Figure 13: Word Cloud View of 'Myeloid Lineage' and Lymphoid Lineage'PubMed Queries of Tumorigenesis Ortholog Subsets. The next query mining focus was myeloid lineage. Within the set of TOP 30% conserved orthologs (upper left), identified terms included inflammation, hypoxia and insulin. The results obtained from the BOTTOM 30% (least conserved orthologs upper right) incude obesity, insulin, infection, vasculitis, polyangiitis, metastasis and the following genes: IL10, IL17, FOxm1, Cxcr4, Cxcl12, cscl10, IL1Ra, CxCL4, GDF15, CXCL1, IL17a, CXCL8. The last queries were focused on lymphoid lineage (Figure 13). The most represented term s associated with the top 30% most conserved orthologs are: inflammation, autoimmune disease, loymphomas, multiple sclerosis, inflammatory disease, and psoriasis. The notable genes were Lef1, mTOR, IL22, Bcl2, IL23, MYD88, IL4, IL6, IRF8, and caspase 1. Within the BOTTOM 30% least conserved genes, the following disorders were identified: multiple sclerosis, arthritis, depression, infection, lymphoma, carcinoma, and infection. The following genes were identified: IL15, CXCL10, IL22, IL21, FoxP3, Cyclin D1, CXCR2, and CD8.

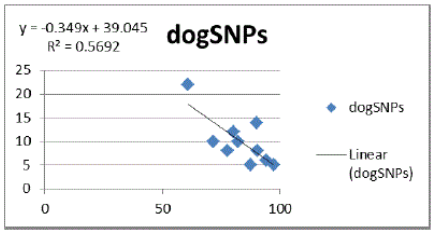

Figure 14A: Scatter Plot of Relationship between Ortholog Identity and Number of SNPs in Dog Genes. The correlation for dog SNPs is shown. For tumorigenesis orthologs having high protein percent identity, approximately 5 SNPs were present in the dbSNP database. As the pairwise percent identity decreased, the number of SNPs increased above 20 for orthologs with very low percent identity.

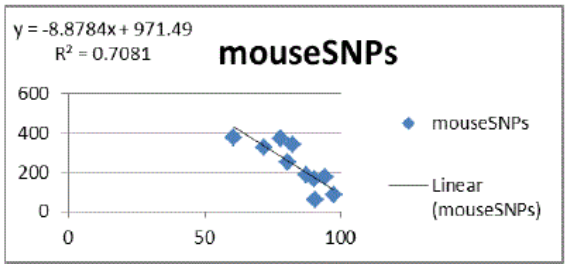

Figure 14B: Scatter Plot of Relationship Between Ortholog Identity and Number of SNPs in Mouse Genes. Scatter plot and correlation for SNPs in mouse genes. Similar to the pattern observed in dogs, the mouse orthologs having the highest percent identity contain the fewest SNPS. For example, mouse orthologs having the highest percent identity have approximately 100 SNPs while the mouse orthologs approaching 60% identity have approximately 400 SNPs.

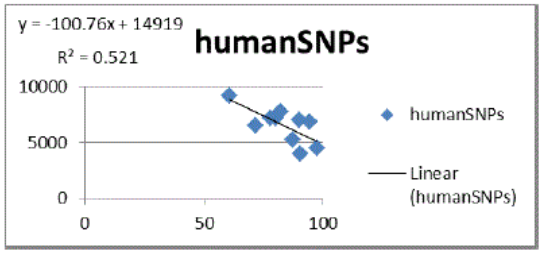

Figure 14C: Scatter Plot of Relationship Between Ortholog Identity and Number of SNPs in Human Genes. The scatter plot for human tumorigenesis orthologs. A similar inverse relationship between ortholog pair-wise percent identity and number of SNPs is observed.

The set of canine tumorigenesis ortholgs was used to query the dbSNP database for protein coding region SNPs. A total of 146 were identified with 33 in the TOP 30% orthologs, 59 in the MIDDLE 40% orthologs and 54 in the BOTTOM 30% orthologs. Within the TOP 30% orthologs 19 missense and 14 frameshift SNPs were identified. The MIDDLE 40% orthologs were associated with 41 missense SNPs, 17 frameshift SNPs and 1 nonsense SNP. The BOTTOM 30% orthologs contained 40 missense SNPs and 14 frameshift SNPs.

Discussion

This comparative genomics approach has provided insight into the biology and evolution of a set of one-to-one orthologous protein coding genes associated with tumorigenesis across human, mouse, dog and naked mole rat. Although humans, mice and dogs are readily susceptible to tumorigenesis, naked mole rats exhibit a resistance to tumorigenesis. Subsequently, the goal of this project was to identify tumorigenesis orthologs in the dog for which pairwise protein identity was most divergent in naked mole rat orthologs, under the hypothesis that such orthologs might represent genes associated with susceptibility and resistance to tumorigenesis in dogs.

Interestingly, the set of most conserved tumorigenesis orthologs, were enriched for developmental processes and organogenesis. Some of these genes were associated with skeleton development and morphology as well as craniofacial development. The production of dog breeds over the last few hundred years has resulted in a considerable range of skeletal and craniofacial morphological phenotypes that are represented by particular breeds. For example, brachycephalic breeds exhibit considerably shorter snouts than either mesencephalic or dolichocephalic breeds. Similarly, dramatic differences in the appendicular skeleton of achondrosplastic versus non achondroplastic dog breeds is very notable. Furthermore, anatomical variation in long bone growth rates, sizes and geometry are easily discerned between dogs of the toy breeds versus dogs of the giant breeds.

The overall conservation observed among the most conserved tumorigenesis orthologs makes sense in the context of their developmental and embryological roles. Many of these genes are associated with embryogenic lethality and premature death. Subsequently, the negative selection acting upon them has maintained their sequence similarity across species as diverse as dog and naked mole rat. This is not to say that artificial selection has not resulted in selection for specific variants of these genes within certain dog breeds, but rather that the negative selection associated with maintaining the developmental programs required for organogenesis has indeed limited the extent of genetic variation within these genes. It is possible that some breed associated tumors may be the result of artificial selection that occurred within these orthologs during the very act of breed formation.

In contrast, the set of least conserved tumorigenesis orthologs are enriched in immunological phenotypes. Immune genes are thought to evolve in response to changing environment and pathogens such as viruses, bacteria as well as parasites [31-35]. Adaptation usually occurs by positive selection or gene duplication and it thought to occur mostly in proteins involved in pathogen recognition and less in molecules involved in cell signaling such as cytokines [36].

Immune responses are tightly regulated to ensure the appropriate balance between inflammatory, anti-inflammatory and regulatory immune cell signaling and dysregulation of any of these elements possibly leads to pathological disease states such as autoimmunity or cancer.

A large number of studies have focused on oncogenes and tumor suppressor genes such as p53, however, recent research suggests that inflammation contributes to cancer development by creating micro-environments favorable for tumorigenesis [37,38]. Several inflammatory conditions such as obesity, Crohn's disease or microbial infections are associated with an increased occurrence of cancer [39-41]. While it has been observed that macrophages are abundant in tumor environments, it was initially thought that these cells infiltrated the tumor in response to tumor growth. In contrast, it appears that macrophages actually mediate the inflammation contributing activated M1 type, which in turn upregulates Th1 signaling leading to the activation of T cells.

Signaling molecules released by M1 type macrophages include TNF and IFNγ associated signaling, IL1b, as well as IL12 and IL23, which direct the differentiation of Th1 and Th17 Tells [42], which in turn further amplify the inflammatory response. In a well-regulated immune response IFNγ signaling actually inhibits tumor growth by stimulating cytotoxic T cells to kill tumor cells. Genes in this group are tightly regulated and dysregulation of any of these proteins potentially facilitates cancer development.

Another cell type involved in preventing tumor development is the NK (natural killer) cell, which recognizes tumor cells as non-self or altered self. Activated NK cells in turn release IL-2 activating T cells as well as IFNγ, activating macrophages. NK cells need to recognize ligands on cells, which bind to its KIR (killer inhibitory receptors), indicating self. An important molecule in the process of self-tolerance and recognition is the molecule TAP (transporter associated with antigen processing), and changes in this molecule might enable tumors to escape recognition by NK cells [43].

On the other hand, tumors create an immunosuppressive microenvironment, involving M2 or anti-inflammatory macrophages and regulatory T cells versus cytotoxic T cells [44-46]. M2 macrophages express different signaling molecules and cell surface receptors than their M1 counterpart, most notably signaling molecules of the TGFb pathway, IL10 and STAT3 signaling and molecules such as VEGF, promoting angiogenesis [47]. Tumor associated macrophages (TAM) recruit regulatory T cells to the tumor environment via chemokine receptors, such as CCR4 and CCR6 as well as the expression of IL-1- [48]. TGFb and IL-10 modulate T cell function creating an immunosuppressive Th2 environment. In addition, the STAT/JAK pathway has been implicated in TGFb mediated establishment of EMT (epithelial-mesenchymal transition), which is considered a key element of early establishment of tumors [49].

The emerging theme is that dysregulation of several inflammatory and anti- inflammatory genes potentially drive tumor development and consequently tumor invasion and metastasis. Several immune cells are involved not only in the detection and destruction of tumors, but can play a role in development and sustaining of cancers by initiating an inflammatory environment in the beginning and switching to an immunosuppressive environment promoting angiogenesis and tumor growth. Several crucial genes in this process show divergence in homology between species and it will be critical to look at these genes in further detail to better understand tumor biology.

Figure 15 displays a set of associations between a subset of canine cancers, dog breeds, and representative members of the conserved and divergent tumorigenesis orthologs that have been independently implicated in osteosarcoma, lymphoma, mammary tumors or mast cell tumors. Breeds associated with osteosarcoma include some giant breeds, such as Irish wolfhound, Great Dane, Saint Bernard, and Afghan hound. The highly conserved tumorigenesis ortholog MEN1 has been implicated in osteosarcoma [6], while four least conserved orthologs have also been implicated in this cancer type: AHR [50], CDN2KB [6], COX2 [51] and PTEN [52].

Breeds associated with lymphoma (Figure 15) include the brachycephalic Bullmastiff, Boxer, and Bulldog as well as the achondroplastic Bassett Hound. Among the most conserved tumorigenesis orthologs, lymphoma is associated with CDKN1A [53], TIAM1 [54] and PTEN [55]. Representative Orthologs from the least conserved set include BRCA1 [53], COX2 [50] and CXCR3 [56].

Another tumor type for which brachycephalic breeds are associated (Boxer, Bulldog, Boston Terrier, and Pug) are mast cell tumors (Figure 15). Three least conserved tumorigenesis orthologs implicated in this tumor type include FOXM1 [57], CCND1 [58] and AHR [50]. In contrast to osteosarcoma, which affects large dog breeds, mammary tumors (Figure 15) tend to occur in the achondroplastic Dachshund and a number of small breeds including the Poodle, Maltese Terrier, Cocker Spaniel, Yorkshire Terrier and the Beagle. Conserved tumorigenesis orthologs implicated in this tumor type include CTNNB1 and PTEN [59] while least conserved orthologs associated with canine mammary tumors include BRCA1 [4] AND [60], BRCA2 [61], MUC1 [62], and KLF4 [63]. Interestingly, tumor associated macrophages (TAMs) have been implicated in modulating tumor invasion and metastasis in this cancer type [64].

The results identified in this analysis suggest that a relationship between breed, morphology and tumorigenesis susceptibility may exist. The high-throughput literature mining that produced the data in Figure 11, identified relationships between 'canine hip dysplasia', 'hip laxity', 'osteoarthritis' and osteosarcoma.' Some of the breeds that are susceptible to osteosarcoma, are also at risk for hip/elbow dysplasia, such as the Saint Bernard [65] as well as the German Shepherd, Labrador Retriever, Golden Retriever and Rottweiler breeds [66]. A genetic association study published in 2012 identified a gene within one of the most conserved tumorigenesis orthologs, FN1, as associated with elbow dysplasia in Bernese Mountain Dogs [67]. Not surprisingly, the FN1 gene is expressed in soft sarcoma and osteosarcoma cells [68].

Craniofacial morphology is, in some cases, a breed defining morphological trait. Evolutionary evidence for the role of the tumorigenesis orthologs in craniofacial variation between species further lends support to a link between breed, morphology and susceptibility to tumorigenesis. Studies in avian species have uncovered multiple developmental mechanisms that give rise to beak shape. The fact that avian evolutionary fitness may depend upon beak morphology in a particular environmental niche suggests craniofacial morphological plasticity might be a desirable developmental program in birds. The tumorigenesis ortholog, beta catenin, expression differs within the premaxillary bone of embryos of species with different beak shapes [69]. Moreover, in some cases, bird species with shared beak shapes exhibit distinct developmental programs underlying beak formation. For example, in one species the beak forms through the action of Bmp4 and Cam followed by TGF beta IIR, Beta-catenin and Dkk3 signaling, while the beak in the other species is formed almost exclusively through the action of Ihh and Bmp4 synergistically cause expansion of bone tissue [70].

The results reported in this comparative genomics analysis are the first to leverage conservation and divergence of canine orthologs of naked mole rat proteins to identify potential relationships between dog breeds, breed associated morphology and tumorigenesis susceptibility. The use of comparative genomics approaches to identify cancer related genes has previously been employed in rats to investigate expansion and contraction of paralogous gene families within the naked mole rat that are implicated in cancer biology [71]. A subsequent naked mole rat comparative genomics approach explored genes associated with genome maintenance in humans and mice [72]. Most recently anti-cancer mechanisms associated with single nucleotide differences identified in naked mole rat proteins has illuminated the aspects of cancer biology in mammals [73].

Additional experimental studies must be carried to validate and further elucidate the role of these tuumorigenesis orthologs in canine morphology, breed formation, health and disease. As with any bioinformatics approach, some of the results will prove to are true positives, while others may end up not proving true. In an attempt to dramatically limit the number of false positive results predicted in this study, we severely limited the number of genes investigated to just a core set of one-to-one orthologs among human, mouse, dog and naked mole rat. Subsequent studies using larger sets of tumorigenesis genes may uncover additional relationships between dog breed, breed morphology and tumorigenesis.

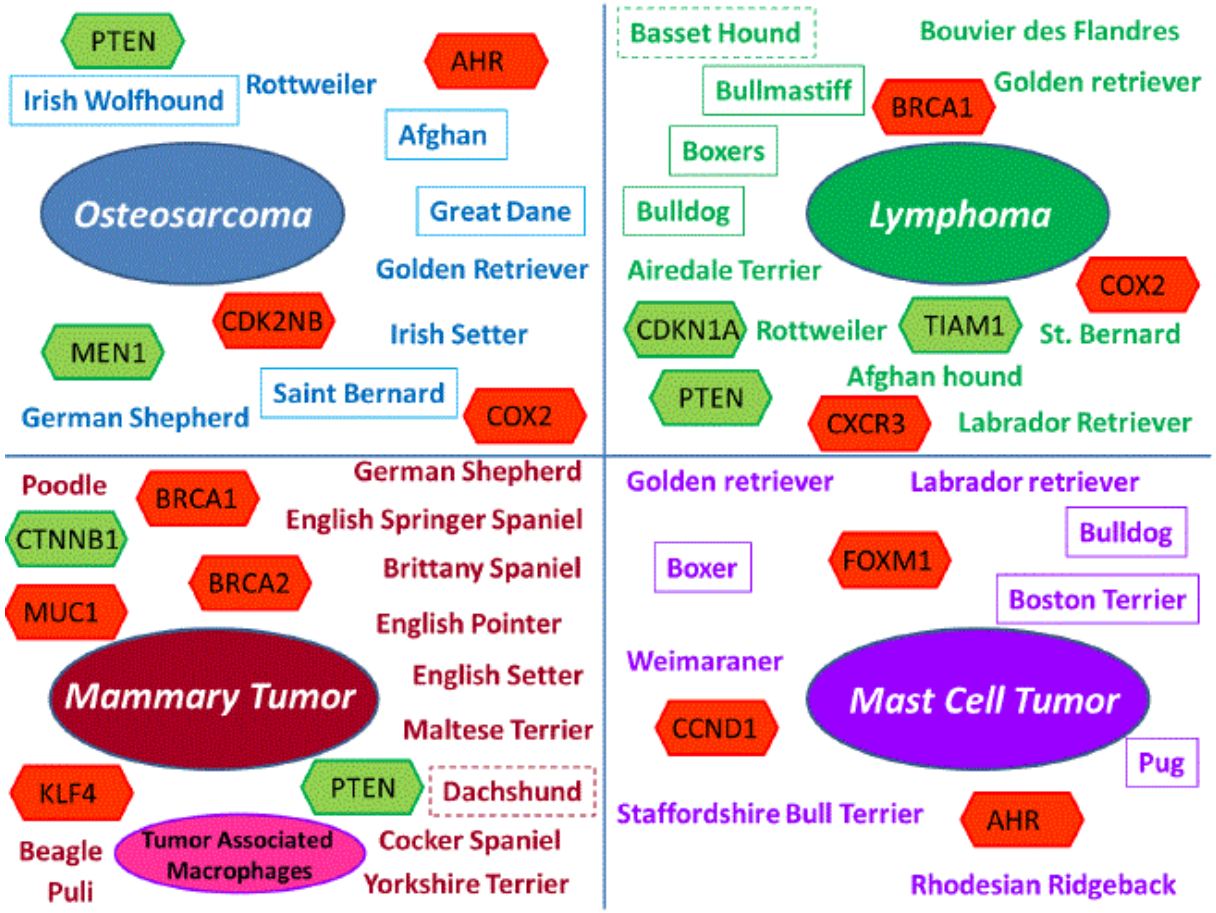

Figure 15: Some Independently Validated Associations of Dog Breed, Canine Tumor Type, and Tumorigenesis Orthologs. Model of a set of associations between canine cancers, dog breeds, and representative members of the conserved and divergent tumorigenesis orthologs. Cancers include osteosarcoma, lymphoma, mammary tumors or mast cell tumors. Breeds associated with osteosarcoma include some giant breeds, such as Irish Wolfhound, Great Dane, Saint Bernard, and Afghan Hound. Breeds associated with lymphoma include the brachycephalic Bullmastiff, Boxer, and Bulldog as well as the achondroplastic Bassett Hound. Another tumor type for which brachycephalic breeds are associated (Boxer, Bulldog, Boston Terrier, and Pug) are mast cell tumors. In contrast to osteosarcoma, which affects large dog breeds, mammary tumors tend to occur in the achondroplastic Dachshund and a number of small breeds including the Poodle, Maltese Terrier, Cocker Spaniel, Yorkshire Terrier and the Beagle. The results illustrated in the figure suggest that a relationship between breed, morphology and tumorigenesis susceptibility may exist.

References

1. Schiffman JD, Breen M (2015) Comparative oncology: what dogs and other species can teach us about humans with cancer. Philos Trans R Soc Lond B Biol Sci 370.
2. Paoloni M, Khanna C (2008) Translation of new cancer treatments from pet dogs to humans. Nat Rev Cancer 8: 147-156.
3. Grüntzig K, Graf R, Hässig M, Welle M, Meier D, et al. (2015) The Swiss Canine Cancer Registry: a retrospective study on the occurrence of tumours in dogs in Switzerland from 1955 to 2008. J Comp Pathol 152: 161-171.
4. Richards KL, Suter SE (2015) Man's best friend: what can pet dogs teach us about non-Hodgkin's lymphoma. Immunol Rev 263: 173-191.
5. Dhawan D, Paoloni M, Shukradas S, Choudhury DR, Craig BA, et al. (2015) Comparative Gene Expression Analyses Identify Luminal and Basal Subtypes of Canine Invasive Urothelial Carcinoma That Mimic Patterns in Human Invasive Bladder Cancer. PLoS One 9: 10.
6. Angstadt AY, Thayanithy V, Subramanian S, Modiano JF, Breen M (2012) A genome-wide approach to comparative oncology: high-resolution oligonucleotide aCGH of canine and human osteosarcoma pinpoints shared microaberrations. Cancer Genet 11: 572-587.
7. Rivera P, Melin M, Biagi T, Fall T, Häggström J, et al. (2009) Mammary tumor development in dogs is associated with BRCA1 and BRCA2. Cancer Res 69: 8770-8774.
8. Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, et al. (2008) Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. Am J Hum Genet 82: 937-948.
9. Soussi T (2014) The TP53 gene network in a postgenomic era. Hum Mutat 35: 641-642.

10. Grochola LF, Zeron-Medina J, Mériaux S, Bond GL (2010) Single-nucleotide polymorphisms in the p53 signaling pathway. Cold Spring Harb Perspect Biol 2: a001032.
11. Setoguchi A, Sakai T, Okuda M, Minehata K, Yazawa M, et al. (2001) Aberrations of the p53 tumor suppressor gene in various tumors in dogs. Am J Vet Res 62: 433-439.
12. Garcia PB, Attardi LD (2014) Illuminating p53 function in cancer with genetically engineered mouse models. Semin Cell Dev Biol 27: 74-85.
13. Deakin JE, Belov K (2012) A comparative genomics approach to understanding transmissible cancer in Tasmanian devils. Annu Rev Genomics Hum Genet 13: 207-222.
14. Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, et al. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA 314: 1850-1860.
15. HrabÄ› de Angelis M, Nicholson G, Selloum M, White JK, Morgan H, et al. (2015) Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. Nat Genet 47: 969-978.
16. Ring N, Meehan TF, Blake A, Brown J, Chen CK, et al. (2015) A mouse informatics platform for phenotypic and translational discovery. Mamm Genome 26: 413-421.
17. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE (2015) The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res 43: D726-36.
18. Liang S, Mele J, Wu Y, Buffenstein R, Hornsby PJ (2010) Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). Aging Cell 9: 626-635.
19. Azpurua J, Seluanov A (2013) Long-lived cancer-resistant rodents as new model species for cancer research. Front Genet 3: 319.
20. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, et al. (2016) Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res 44: D574-580.
21. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, et al. (2015) Gene: a gene-centered information resource at NCBI. Nucleic Acids Res 43: D36-42.
22. Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24: 319-324.
23. Becker KG, Hosack DA, Dennis G Jr, Lempicki RA, Bright TJ, et al. (2003) PubMatrix: a tool for multiplex literature mining. BMC Bioinformatics 4: 61.
24. David J, Irizarry KJ (2009) Using the PubMatrix literature-mining resource to accelerate student-centered learning in a veterinary problem-based learning curriculum. J Vet Med Educ 36: 202-208.
25. Parker DS, Chu WW, Sabb FW, Toga AW, Bilder RM (2009) Literature Mapping with PubAtlas - extending PubMed with a 'BLASTing interface'. Summit on Translat Bioinforma 2009: 90-94.
26. Wei CH, Harris BR, Li D, Berardini TZ, Huala E, et al. (2012) Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. Database pp: 17.
27. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48.
28. Weng MP, Liao BY (2010) MamPhEA: a web tool for mammalian phenotype enrichment analysis. Bioinformatics 26: 2212-2213.
29. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, et al. (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 28: 1805-1806.
30. Bhagwat M (2010) Searching NCBI's dbSNP database. Curr Protoc Bioinformatics Chapter 1: Unit 1.
31. Schmid-Hempel P (2003) Variation in immune defence as a question of evolutionary ecology. Proc Biol Sci 270: 357-366.
32. Rothenburg S, Seo EJ, Gibbs JS, Dever TE, Dittmar K (2009) Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. Nat Struct Mol Biol 16: 63-70.
33. Webb AE, Gerek ZN, Morgan CC, Walsh TA, Loscher CE, et al. (2015) Adaptive Evolution as a Predictor of Species-Specific Innate Immune Response. Mol Biol Evol 32: 1717-1729.
34. Barribeau SM, Sadd BM, du Plessis L, Schmid-Hempel P (2014) Gene expression differences underlying genotype-by-genotype specificity in a host-parasite system. Proc Natl Acad Sci U S A 111: 3496-3501.
35. Carattoli A, Seiffert SN, Schwendener S, Perreten V, Endimiani A (2015) Differentiation of IncL and IncM Plasmids Associated with the Spread of Clinically Relevant Antimicrobial Resistance. PLoS One 10: e0123063.
36. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, et al. (2007) Dynamic evolution of the innate immune system in Drosophila. Nat Genet 39: 1461-1468.
37. Joyce JA, Pollard JW (2009) Microenvironmental regulation of metastasis. Nat Rev Cancer 9: 239-252.
38. Balkwill FR, Mantovani A (2012) Cancer-related inflammation: common themes and therapeutic opportunities. Semin Cancer Biol 22: 33-40.
39. Grivennikov SI, Greten FR, Karin M (2010) Immunity, inflammation, and cancer. Cell 140: 883-899.
40. Howe LR, Subbaramaiah K, Hudis CA, Dannenberg AJ (2013) Molecular pathways: adipose inflammation as a mediator of obesity-associated cancer. Clin Cancer Res 19: 6074-6083.
41. Balkwill F, Charles KA, Mantovani A (2005) Smoldering and polarized inflammation in the initiation and promotion of malignant disease. Cancer Cell 7: 211-217.
42. Omrane I, Baroudi O, Bougatef K, Mezlini A, Abidi A, et al. (2014) Significant association between IL23R and IL17F polymorphisms and clinical features of colorectal cancer. Immunol Lett 158: 189-194.
43. Hanada K, Yewdell JW, Yang JC (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. Nature 427: 252-256.
44. Coussens LM, Pollard JW (2011) Leukocytes in mammary development and cancer. Cold Spring Harb Perspect Biol 3.
45. Gajewski TF, Schreiber H, Fu YX (2013) Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 14: 1014-1022.
46. Pollard JW (2004) Tumour-educated macrophages promote tumour progression and metastasis. Nat Rev Cancer 4: 71-78.
47. Noy R, Pollard JW (2014) Tumor-associated macrophages: from mechanisms to therapy. Immunity 41: 49-61.
48. Savage ND, de Boer T, Walburg KV, Joosten SA, van Meijgaarden K (2008) et al. Human anti- inflammatory macrophages induce Foxp3+ GITR+ CD25+ regulatory T cells, which suppress via membrane-bound TGFbeta-1. J Immunol 181: 2220-2206.
49. Liu L, Chen X, Wang Y, Qu Z, Lu Q, et al. (2014) Notch3 is important for TGF-Î²-induced epithelial-mesenchymal transition in non-small cell lung cancer bone metastasis by regulating ZEB-1. Cancer Gene Ther 21: 364-372.
50. Giantin M, Vascellari M, Lopparelli RM, Ariani P, Vercelli A, et al. (2013) Expression of the aryl hydrocarbon receptor pathway and cyclooxygenase-2 in dog tumors. Res Vet Sci 94: 90-99.
51. Millanta F, Asproni P, Cancedda S, Vignoli M, Bacci B, et al. (2012) Immunohistochemical expression of COX-2, mPGES and EP2 receptor in normal and reactive canine bone and in canine osteosarcoma. J Comp Pathol 147: 153-160.
52. Levine RA, Forest T, Smith C (2002) Tumor suppressor PTEN is mutated in canine osteosarcoma cell lines and tumors. Vet Pathol 39: 372-378.
53. Zamani-Ahmadmahmudi M, Najafi A, Nassiri SM (2015) Reconstruction of canine diffuse large B-cell lymphoma gene regulatory network: detection of functional modules and hub genes. J Comp Pathol 152: 119-130.
54. Shepherd TR, Klaus SM, Liu X, Ramaswamy S, DeMali KA, et al. (2010) The Tiam1 PDZ domain couples to Syndecan1 and promotes cell-matrix adhesion. J Mol Biol 398: 730-746.

55. Elvers I, Turner-Maier J, Swofford R, Koltookian M, Johnson J, et al (2015) Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. Genome Res 25: 1634-1645.
56. Chimura N, Iio A, Ozaki E, Mori T, Ito Y, et al. (2013) Transcription profile of chemokine receptors, cytokines and cytotoxic markers in peripheral blood of dogs with epitheliotropic cutaneous lymphoma. Vet Dermatol 24: 628-631.
57. Giantin M, Granato A, Baratto C, Marconato L, Vascellari M, et al. (2014) Global gene expression analysis of canine cutaneous mast cell tumor: could molecular profiling be useful for subtype classification and prognostication? PLoS One 9: e95481.
58. Ozaki K, Yamagami T, Nomura K, Narama I (2007) Prognostic significance of surgical margin, Ki-67 and cyclin D1 protein expression in grade II canine cutaneous mast cell tumor. J Vet Med Sci 69: 1117-1121.
59. Uva P, Aurisicchio L, Watters J, Loboda A, Kulkarni A, et al. (2009) Comparative expression pathway analysis of human and canine mammary tumors. BMC Genomics 10: 135.
60. Flisikowski K, Flisikowska T, Sikorska A, Perkowska A, Kind A, et al. (2015) Germline gene polymorphisms predisposing domestic mammals to carcinogenesis. Vet Comp Oncol.
61. Yoshikawa Y, Morimatsu M, Ochiai K, Ishiguro-Oonuma T, Wada S, et al. (2015) Reduced canine BRCA2 expression levels in mammary gland tumors. BMC Vet Res 11: 159.
62. Campos LC, Silva JO, Santos FS, Araújo MR, Lavalle GE, et al. (2015) Prognostic significance of tissue and serum HER2 and MUC1 in canine mammary cancer. J Vet Diagn Invest 27: 531-535.
63. Tien YT, Chang MH, Chu PY, Lin CS, Liu CH, et al. (2015) Downregulation of the KLF4 transcription factor inhibits the proliferation and migration of canine mammary tumor cells. Vet J 205: 244-253.
64. Król M, Mucha J, Majchrzak K, Homa A, Bulkowska M, et al. (2014) Macrophages mediate a switch between canonical and non-canonical Wnt pathways in canine mammary tumors. PLoS One 9: e83995.
65. Nelson LL, Dyce J, Shott S (2007) Risk factors for ventral luxation in canine total hip replacement. Vet Surg 36: 644-653.
66. Soo M, Sneddon NW, Lopez-Villalobos N, Worth AJ (2015) Genetic evaluation of the total hip score of four populous breeds of dog, as recorded by the New Zealand Veterinary Association Hip Dysplasia Scheme (1991-2011). NZ Vet J 63: 79-85.
67. Pfahler S, Distl O (2012) Identification of quantitative trait loci (QTL) for canine hip dysplasia and canine elbow dysplasia in Bernese mountain dogs. PLoS One 7: e49782.
68. Bai C, Yang M, Fan Z, Li S, Gao T, et al. (2015) Associations of chemo- and radio-resistant phenotypes with the gap junction, adhesion and extracellular matrix in a three-dimensional culture model of soft sarcoma. J Exp Clin Cancer Res 10: 34:58.
69. Mallarino R, Grant PR, Grant BR, Herrel A, Kuo WP, et al. (2011) Two developmental modules establish 3D beak-shape variation in Darwin's finches. Proc Natl Acad Sci USA 108: 4057-4062.
70. Mallarino R, Campàs O, Fritz JA, Burns KJ, Weeks OG, et al (2012) Closely related bird species demonstrate flexibility between beak morphology and underlying developmental programs. Proc Natl Acad Sci USA 2: 16222-16227.
71. Yang Z, Zhang Y, Chen L (2013) Investigation of anti-cancer mechanisms by comparative analysis of naked mole rat and rat. BMC Syst Biol 7 Suppl 2: S5.
72. MacRae SL, Zhang Q, Lemetre C, Seim I, Calder RB, et al. (2015) Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. Aging Cell 14: 288-291.
73. Yang Z, Zhang Y, Chen L (2015) Single amino acid changes in naked mole rat may reveal new anti-cancer mechanisms in mammals. Gene 572: 101-107.