UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1454

Analysis of inherited and somatic variants to decipher canine complex traits

KATE MEGQUIER

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6206
ISBN 978-91-513-0310-9
urn:nbn:se:uu:diva-347165

Dissertation presented at Uppsala University to be publicly examined in B:22, BMC, Husargatan 3, Uppsala, Monday, 21 May 2018 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: David Sargan (University of Cambridge).

Abstract

Megquier, K. 2018. Analysis of inherited and somatic variants to decipher canine complex traits. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1454. 67 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0310-9.

This thesis presents several investigations of the dog as a model for complex diseases, focusing on cancers and the effect of genetic risk factors on clinical presentation.

In Papers I and II, we performed genome-wide association studies (GWAS) to identify germline risk factors predisposing US golden retrievers to hemangiosarcoma (HSA) and B-cell lymphoma (BLSA). Paper I identified two loci predisposing to both HSA and BLSA, approximately 4 megabases (Mb) apart on chromosome 5. Carrying the risk haplotype at these loci was associated with separate changes in gene expression, both relating to T-cell activation and proliferation.

Paper II followed up on the HSA GWAS by performing a meta-analysis with additional cases and controls. This confirmed three previously reported GWAS loci for HSA and revealed three new loci, the most significant on chromosome 18. This locus contains several candidate genes with a clear role in carcinogenesis, including *KMT5B* and *LRP5*. Overall, carriers of the risk alleles at the top six loci are diagnosed with HSA earlier in life.

In Paper III we investigated the somatic mutations which occur in HSA tumor tissue by performing tumor-normal exome sequencing of 47 golden retrievers. We identified 7 recurrently mutated genes, including the tumor suppressor *TP53* (mutated in 59.6% of tumors) and oncogene *PIK3CA* (mutated in 29.8% of tumors). Additional somatically mutated genes overlap those found in human angiosarcomas, suggesting that angiosarcomas in dogs and humans are genetically very similar.

In Paper IV, we investigated the variable penetrance of a *SOD1* mutation in Pembroke Welsh corgis causing degenerative myelopathy (DM), a model of the human motor neuron disease amyotrophic lateral sclerosis (ALS). We discovered that regulatory variants near the *SP110* gene were associated with an increased risk of DM and an earlier age at diagnosis, suggesting a role for immune response in the pathogenesis of the disease.

Taken together, these findings provide new insight into the pathophysiology of both hemangiosarcoma and degenerative myelopathy, which could guide future diagnostics and therapeutic strategies both in humans and veterinary patients. In addition, they demonstrate the power of the dog as a biomedical model for human complex diseases.

Keywords: dog, genetics, GWAS, exome, cancer, DM

Kate Megquier, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-0310-9
urn:nbn:se:uu:diva-347165 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-347165)

Between animal and human medicine, there is no dividing line –
nor should there be. The object is different, but the experience obtained
constitutes the basis of all medicine.
- Rudolf Ludwig Carl Virchow, 1856

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Tonomura N, Elvers I, Thomas R, **Megquier K**, Turner-Maier J, Howald C, Sarver AL, Swofford R, Frantz AM, Ito D, Mauceli E, Arendt M, Noh HJ, Koltookian M, Biagi T, Fryc S, Williams C, Avery AC, Kim JH, Barber L, Burgess K, Lander ES, Karlsson EK, Azuma C, Modiano JF*, Breen M*, Lindblad-Toh K*. *Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers.* PLoS Genet. 2015 Feb;11(2):e1004922. **Authors contributed equally.*

II **Megquier K**, Tonomura N, Fall T, Noh HJ, Turner-Maier J, Swofford R, Koltookian M, Biagi T, Fryc S, Arendt M, Häggström J, Barber L, Burgess K, Thomas R, Breen M, Modiano J, Elvers I, Azuma C, Karlsson E*, Lindblad-Toh K*. *Genome-wide meta-analysis identifies inherited variation contributing to overall risk and age of onset in canine angiosarcoma.* Manuscript. 2018. **Authors contributed equally.*

III **Megquier K**, Turner-Maier J, Swofford R, Kim JH, Sarver A, Wang C, Sakthikumar S, Johnson J, Koltookian M, Lewellen M, Scott M, Graef A, Tonomura N, Fryc S, Biagi T, Alfoldi J, Thomas R, Karlsson E, Breen M*, Modiano J*, Elvers I*, Lindblad-Toh K*. *Exome sequencing of hemangiosarcomas in the golden retriever reveals frequent mutation of TP53 tumor suppressor and PIK3CA oncogene.* Manuscript. 2018. **Authors contributed equally.*

IV Ivansson EL, **Megquier K**, Kozyrev SV, Murén E, Körberg IB, Swofford R, Koltookian M, Tonomura N, Zeng R, Kolicheski AL, Hansen L, Katz ML, Johnson GC, Johnson GS, Coates JR, Lindblad-Toh K. *Variants within the* SP110 *nuclear body protein modify risk of canine degenerative myelopathy*. Proc Natl Acad Sci U S A. 2016 May 31;113(22):E3091-100.

Reprints were made with permission from the respective publishers.

Contents

Introduction....................11
Complex disease in humans and dogs....................12
Genetic analysis of complex disease....................12
The dog as a model for human disease....................13
Cancer....................14
Inherited risk factors....................14
Somatic driver mutations....................15
Cancer therapy....................16
Overview of cancer model organisms....................16
Canine model of cancer....................17
Neurodegenerative diseases....................19
Canine model of neurodegenerative disease....................20
Genetic mapping of complex traits in the dog....................21
Methods....................22
Genome-wide association studies (GWAS)....................22
Whole-genome imputation....................24
Pathway analysis....................25
Whole exome sequencing....................26
Sequencing of somatic variants....................27
Functional evaluation of genetic variants....................29
Clinical variables....................31
Canine hemangiosarcoma....................31
Biology of HSA and comparative aspects with angiosarcoma (AS)....................31
Genetic studies of human AS and canine HSA....................32
Canine degenerative myelopathy....................34
Biology of DM and comparative perspective with ALS....................34
Genetic studies of human ALS and canine DM....................34

Aims of the thesis....................36

Summary of included papers 37
Part 1 - Germline risk factors in hemangiosarcoma 37
Background...37
Paper I: Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers 37
Paper II: Genome-wide meta-analysis identifies inherited variation contributing to overall risk and age of onset in canine angiosarcoma 38
Discussion..39
Part 2 – Somatic mutations in hemangiosarcoma 40
Background...40
Paper III - Exome sequencing of hemangiosarcomas in the golden retriever reveals frequent mutation of the *TP53* tumor suppressor and *PIK3CA* oncogene..40
Discussion..42
Part 3 - Modification of risk in canine degenerative myelopathy with *SOD1* mutation 43
Background...43
Paper IV - Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy 43
Discussion..44

General discussion 45

Future directions 48

Acknowledgments...50

References...52

Abbreviations

ALS	amyotrophic lateral sclerosis
ARPC1A	actin related protein 2/3 complex subunit 1A
ATP5H	ATP synthase peripheral stalk subunit d (ATP5PD)
AS	angiosarcoma
BLSA	B-cell lymphoma
CDK2	cyclin-dependent kinase 2
CDK2AP2	cyclin-dependent kinase 2 associated protein 2
CDKN2A/B	cyclin-dependent kinase inhibitor 2A/B
CNV	copy number variant
DM	degenerative myelopathy
EGFR	epidermal growth factor receptor
ETA	endothelin receptor type A (*EDNRA*)
FGF2	fibroblast growth factor 2
Flt-1	fms related tyrosine kinase 1 (*VEGFR1*)
Flk-1	kinase insert domain receptor (*KDR*, *VEGFR2*)
GR	golden retriever
GWAS	genome-wide association study
HSA	hemangiosarcoma
KDM2A	lysine (K)-specific demethylase 2A
KMT5B	lysine (K)-specific methyltransferase 5B
KRAS	KRAS proto-oncogene, GTPase
LSA	lymphoma
LD	linkage disequilibrium
NRAS	NRAS proto-oncogene, GTPase
OR	odds ratio
ORC1	origin recognition complex subunit 1
OSA	osteosarcoma
PDX	patient-derived xenograft
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PIK3R1	phosphoinositide-3-kinase regulatory subunit 1
PPET1	endothelin 1 (*EDN1*)
PTEN	phosphatase and tensin homolog
PWC	Pembroke Welsh corgi

RASA1	RAS p21 protein activator 1
RB1	retinoblastoma 1
SCNA	somatic copy number aberration
SH3GL2	SH3 domain containing GRB2 like 2
SNP	single nucleotide polymorphism
SOD1	superoxide dismutase 1
SP110	SP110 nuclear body protein
TRPC6	transient receptor potential cation channel, subfamily C, member 6
VEGF	vascular endothelial growth factor
WES	whole exome sequencing
WGS	whole genome sequencing

Introduction

Pet dogs have the potential to be an exceptional model for human complex diseases, and to be man's best friend in ushering in the era of precision medicine. Our canine companions suffer from many of the same diseases as their human owners, including cancers, such as hemangiosarcoma (HSA) (angiosarcoma (AS) in humans), osteosarcoma (OSA), and non-Hodgkin lymphoma (LSA); neurodegenerative disorders such as amyotrophic lateral sclerosis (ALS); immune disorders such as systemic lupus erythematosus (SLE) and atopic dermatitis (AD); and many more. Indeed, the list of complex diseases that we have in common with our canine companions is nearly as long as the list of complex diseases itself, which has long begged the question of whether dogs have a role to play as biomedical models, or perhaps partners in the effort to understand and cure these devastating diseases in all species.

Complex diseases are disorders characterized by polygenic inheritance and interaction with environmental risk factors[1]. They include some of the biggest health concerns our society faces today, including cancers, psychiatric disorders, immune disorders, and neurodegenerative diseases. New diagnostics and therapeutic strategies are urgently needed. However, due to their polygenic nature and the complexity of tracking environmental exposures, dissecting the genetic underpinnings of complex diseases has proven extremely difficult[1]. Comparative genomics using model organisms, where these diseases are less complex, offers a path forward.

Dogs offer the ability to study the genetics of these complex diseases with a simplified genetic background within breeds; they present similar advantages to human population isolates[2]. Due to their unique population history, having gone through one population bottleneck at domestication, and another more recent bottleneck during breed creation, dogs have much longer linkage disequilibrium (LD) and are more homogenous within breeds than human populations[2,3]. There has been a long history of the exchange of comparative medical findings between canine and human medicine, and *vice versa*[4]. For example, while many of the chemotherapy protocols pet dogs receive are adapted from human protocols, in some cases further testing in veterinary patients has refined chemotherapy protocols in humans[4]. In addition, the limb-sparing surgery used in osteosarcoma treatment in both species was pioneered in the dog[4,5].

This thesis is an exploration of the power of pet dogs to facilitate the genetic dissection of complex diseases of importance in both human and veterinary medicine. It begins with three studies providing an in-depth exploration of germline and somatic mutations in hemangiosarcoma, a common cancer in dogs which may model the rare and aggressive cancer angiosarcoma in humans. The final study describes the discovery of a modifier of risk in canine degenerative myelopathy, a model for *SOD1*-mutant ALS in humans.

I will begin by presenting background information on the topics and techniques discussed in these studies, including concepts such as complex trait genetics in humans and dogs, background and comparative information on cancer and neurodegenerative diseases in dogs and humans, methods such as genome-wide association studies, whole-genome imputation, whole exome sequencing, somatic variant calling, and various functional assays.

Complex disease in humans and dogs

The term "complex disease" can be used in different contexts, but here refers to diseases with a polygenic architecture, rather than monogenic or Mendelian inheritance patterns, that often involve some interaction with the environment, which modifies risk of disease[1,6]. Complex diseases include many of the critical health issues facing our society today, including cancers, psychiatric disorders, cardiovascular disease, and diabetes mellitus, with a correspondingly large focus from the biomedical community on new diagnostics and therapeutics for these diseases. However, it is much more difficult to uncover the underlying genetic basis of complex disease than Mendelian disease, as tens, hundreds or even thousands of loci may contribute to risk of a given complex disease, with each individual locus only contributing a small fraction of the total risk (odds ratios typically under 1.1 - 1.3)[7]. In addition, genetic heterogeneity within and between populations further complicates the identification of genetic risk factors, as different combinations of risk factors are carried by different patients, and no one risk factor is necessary to cause disease[8,9].

Genetic analysis of complex disease

In the post-genomic era, we have the ability to study the underlying genetics of complex diseases at unprecedented scale and depth. The most common technique used to discover new associations to complex diseases and traits has been the genome-wide association study (GWAS, discussed in greater detail below)[7,10]. GWAS allows mapping of risk loci in large unrelated cohorts of cases and controls, without *a priori* knowledge of the mechanisms of disease pathogenesis[8,10,11]. As of March, 2018, the NHGRI-EBI catalog of published GWAS findings included 3,329 publications reporting 59,707 associations[12]. However, it has become clear that larger and larger datasets will be needed in order to map risk variants with small effect sizes[7,10]. Indeed, despite the accumulation of massive datasets through large research consortia, which have led to the discovery of many new risk loci, we will never uncover all risk loci associated with all diseases using current methods in human genetics. Visscher, *et al.* performed power calculations demonstrating that detection of associations of some variant frequencies and effect sizes would require a sample size approaching or exceeding the entire human population of the Earth[13]. This is where a comparative approach could prove invaluable.

The dog as a model for human disease

Pet dogs have many advantages as a model organism. Many of the diseases that they share with their human owners arise as complex, polygenic traits, allowing interrogation of the underlying genetics as they naturally arise, something that is less feasible in mouse models where disease often is induced via single gene knockout. The initial publication of the canine genome in 2005 showed that the dog genome is more similar to the human genome than the mouse genome is, in terms of conserved ancestral sequence, sequence composition, and rate of nucleotide divergence[3,14].

Dogs receive a high level of veterinary care and disease monitoring[15], as over 60% of US owners consider the dog to be a member of the family[16]. In addition, dogs age more rapidly than humans, enabling faster data acquisition and conclusion of clinical trials[17–20].

Our dogs also share our environment to a large degree - an important point, as the complex role of environmental exposures in disease pathophysiology is often difficult to assess. Dog disease surveillance can be an early warning for human disease risk at the population level, and can highlight possible disease-exposure links that may not be clear from human studies alone[21]. Many pet dogs spend their entire lives in the same shared environment with their owners, and detailed information can be collected on both exposures in the home and medical history[21,22]. Information on these disease-exposure links can be invaluable, as there are many examples linking both canine and human cancers to the same environmental exposure[21,22]. Environmental tobacco smoke, one of the best-known risk factors in human respiratory tract cancers, has been shown to increase risk of nasal and sinus cancer[23] and lung cancer[24] in dogs. Dogs diagnosed with mesothelioma were more likely to have an owner who was occupationally exposed to asbestos, a known risk factor for mesothelioma in humans[25]. Dogs living in counties with a higher level of industrial activity were found to have a higher rate of bladder cancer, parallelling human mortality data[26]. A weak association was found between phenoxyacid herbicides and risk of lymphoma in dogs[27], while herbicide exposure has been hypothesized to contribute to the elevated risk of non-Hodgkin's lymphoma among farmers. Military dogs who served in the Vietnam War were found to be at increased risk of testicular cancer[28], supporting studies that indicated a possible increased risk in soldiers who had served in Vietnam, as well[29].

Cancer

Cancer is a complex and heterogeneous family of diseases characterized by abnormal, uncontrolled cell replication, invasiveness, and/or spread. Cancer occurs across the animal kingdom, from humans, to companion animals, non-mammalian vertebrates, and even invertebrates[30–32]. Cancers are typically first classified by the organ and cell type of origin, with the major subtypes being carcinomas (epithelial origin), sarcomas (mesenchymal origin), leukemias (hematopoietic cell origin), lymphomas and myelomas (immune cell origin), and central nervous system cancers[30].

Cancers are truly diseases of the genome[33], with both inherited and acquired genetic variation playing important roles. In order to become a cancer, a cell must gain a number of abilities that normal cells lack, including (as outlined by Hanahan and Weinberg in their seminal 2000 paper "Hallmarks of Cancer"), self-sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, limitless growth potential, sustained angiogenesis, and evasion of apoptosis[34]. In their follow-up paper in 2011, Hanahan and Weinberg added the "Emerging Hallmarks" of immune evasion and deregulating cellular energetics, as well as the "Enabling Characteristics" of genome instability and mutation, and tumor-promoting inflammation[35].

Inherited risk factors

Germline risk factors behind the Mendelian cancer syndromes have been easier to discover, as they tend to be inherited in an autosomal dominant fashion. These syndromes include Li-Fraumeni syndrome, in which inherited mutations often in the *TP53* tumor suppressor gene lead to a greatly increased risk of multiple cancers[36]. Patients with inherited mutations in the *RB1* tumor suppressor gene are highly predisposed to developing retinoblastomas in childhood[37]. Statistical analysis of age of retinoblastoma onset and *RB1* mutational status led to the development of Knudson's well-known "two-hit hypothesis[37]." The "two hit hypothesis" explains that in order for cancer to occur, both copies of a tumor suppressor must be deactivated in a given cell; children who inherited a defective copy at birth only needed one additional "hit" to knock out the second copy, thus they developed cancer at a younger age[37]. Inherited defects in the *BRCA1* or *BRCA2* tumor suppressor genes cause perhaps one of the most well-known cancer predisposition syndromes, leading to increased risk of breast, ovarian, and prostate cancers[38,39]. The BRCA genes play a role in maintaining genomic stability through double-stranded DNA break (DSB) repair[40]. However, even Mendelian-inherited cancer syndromes can be incompletely penetrant and depend on both modifier genes and environmental factors[41]. The majority of cancer cases are sporadic, rather than Mendelian.

Somatic driver mutations

The other type of mutation that is important in cancer is somatic mutation, a mutation arising in a somatic cell that is not inherited and will not be passed on to the offspring[42]. DNA damage is most commonly caused by endogenous processes, such as hydrolysis of the phosphodiester bond leading to depurination, oxidation from reactive oxygen species, and alkylation[43,44]. Damage can also arise through exposure to carcinogens such as ultraviolet light in skin cancers[45] or tobacco smoke in lung carcinomas[46]. These errors become mutations when the DNA repair machinery fails to catch them, and the DNA is replicated with the defect in place[43,47]. Damage will accumulate more quickly in cells with defects in DNA repair genes, although this is not necessary for oncogenesis[48]. Types of mutations can include single nucleotide variants (SNVs); insertions or deletions of various sizes (indels); somatic copy number aberrations (SCNAs), which can include amplifications or deletions; or rearrangements, which can result in gene fusions[42].

Not all mutations that accumulate in a cancer cell necessarily have an effect on their phenotype. Stratton, *et al.* defined "driver mutations" as those which provide the cancer cell with a selective advantage over other cells, and "passenger mutations" as those that do not confer selective advantage[42]. Distinguishing between the two types of somatic mutations in order to identify potential therapeutic targets is one of the main challenges in cancer genomics[42].

Two types of genes that are commonly affected (as germline risk factors or somatic driver mutations) are oncogenes and tumor suppressors[30]. Often oncogenes act in a dominant manner, and are constitutively activated or amplified in the tumor, providing signals to grow and divide, while tumor suppressor genes (like *RB1* and *BRCA1*) act in a recessive manner, and are deactivated by mutations or deleted[30]. One metaphor that is used commonly is that, if the cancer cell is a car, an oncogene is like the gas pedal[49], providing the "go" signal (or the "self-sufficiency in growth signaling" hallmark, while the tumor suppressor is like the brakes[49], providing the "stop" signal, and disabling it provides the "insensitivity to anti-growth signals" hallmark[34]. Mutations in either of these types of genes can be drivers of oncogenesis.

Cancer therapy

Cancer patients often do not die of their primary tumor, but rather from metastases, or the spread of the tumor to other locations in the body[50]. The mainstays of clinical cancer therapy include surgical resection of the tumor if possible, followed by radiation therapy or chemotherapy to kill microscopic amounts of remaining tumor cells[31]. Most chemotherapeutic agents are nonspecific - they target any fast-dividing cell[51]. This is why chemotherapy often causes patients to become prone to infection or to feel nauseous - their normal immune and gastrointestinal cells are also being destroyed. In the past decade, however, the discovery of driver mutations for certain types of cancer has led to the advent of "targeted therapeutics[52]." These drugs can target cancer cells with a particular mutation, while sparing the normal tissue[52]. This is one of the goals of precision medicine - to be able to sequence a patient's tumor and select a targeted therapy specifically designed for the mutations in that tumor, without causing damage to the normal tissue.

Overview of cancer model organisms

Human cancers are complex polygenic diseases influenced by environmental factors[30,31,42]. Traditionally, murine models have been used to study cancers in vivo, using single-gene knockouts, tissue xenograft, or exposure to known carcinogens[53,54]. The strengths of mouse model include the existence of many well-characterized inbred lines and the ability to completely control their environment. However, these mouse models may not adequately capture the complexity of human cancers, which have a long growth period, complex underlying germline and somatic genome alterations, are heterogeneous within and between individuals, and have complex interaction with the tumor microenvironment, immune system, and any therapeutics administered[53,55]. Immunodeficient mice have been used as hosts for human cancer cell lines or primary tumor tissue[56]. Newer, IL-2 deficient mice can be engrafted with patient-derived xenograft from solid or hematological cancers[56]. These mice can also be engrafted with human immune cells[56,57]. In recent years, "humanized mice," which are immunodeficient mice engrafted with cells of human origin

- for example, hematopoietic bone marrow cells, pancreatic islet cells, liver cells, and different cancers - have been developed in order to better mimic the responses of human cell types[57]. These humanized patient-derived xenograft (PDX) mice can to some extent recapitulate the complex interaction of the immune system with the tumor, and, it is hoped, will be able to be an “avatar” for the human patient[56], an approach which has shown some success in the clinic[58].

Sometimes it is the differences, rather than the similarities that are of interest in comparative medicine[59]. For example, the naked mole rat and blind mole rat are both highly resistant to neoplasia and live extended lifespans given their body sizes. It is thought that this may be due to adaptation to their environment, in which they must be resistant to hypoxic conditions, as well as heavy metals and other toxic compounds found in their subterranean environment. Blind mole rats were found to repress excessive cellular replication with interferon-beta mediated coordinated cell death[60], while naked mole rat cells display early contact inhibition of dividing cells, mediated by high-molecular-mass hyaluronan[61].

“Peto’s paradox” refers to the observation that, while animals with larger numbers of cells should theoretically have a higher cancer risk, if we look across many different species, larger animals do not tend to have higher cancer rates[62]. This suggests that larger organisms have evolved mechanisms to prevent cancer, particularly given that, within a species, size does correlate with cancer risk (*e.g.* in humans as well as dogs)[62]. For example, the elephant genome has 12 orthologs of the tumor suppressor *TP53*, which, if functional, may account for reduced cancer risk[62].

Canine model of cancer

Pet dogs suffer from many of the same cancers as humans, including osteosarcoma, non-Hodgkin’s lymphoma, mast cell tumors, melanoma, mammary carcinoma, and hemangiosarcoma (angiosarcoma), among others[4,17,20,31,53,63–67]. Cancer is the leading cause of death in pet dogs[68]. Canine cancers are often clinically and pathologically similar to human cancers[4,17,20,53,63–67].

The risk for different cancers in dogs segregates by breed, as well as with certain morphologic traits, suggesting that risk is strongly influenced by genetic background. Some examples of breed-specific risk include the increased risk of HSA and LSA in the golden retriever[69], increased risk of histiocytic sarcoma in Bernese mountain dogs[68,70,71], and risk of transitional cell carcinoma of the bladder in Scottish terriers[68,70]. Examples of cancer risk which segregates among different breeds with similar morphological traits include the increased risk of osteosarcoma in larger, long-legged breeds, such as greyhounds, Rottweilers, Irish wolfhounds, and great Danes[68,70,72]; and the increased risk of gliomas in brachycephalic breeds such as boxers, French and English bulldogs, and Boston terriers[70,73].

In addition, although humans and dogs are susceptible to many of the same cancers, the relative frequency of the cancers between the two species is quite different. For example, the risk of osteosarcoma in dogs is 13.9/100,000[19,68], while in humans it is 10 times less - 1.02/100,000[19,74]. In a general population survey, hemangiosarcoma accounted for 5% of canine tumors[31], while in humans, only about 1% of all cancers are sarcomas[75], and 2-3% of these are angiosarcomas[76]. Some cancers that are rare in humans are very common in dogs, and thus may enable larger genetic studies that would not be feasible in humans.

Lymphoma (LSA) is a cancer of lymphocytes, a type of white blood cell[31]. It is the most common hematologic cancer in the dog, accounting for 15% of all canine malignancies and 79% of all hematopoietic and lymphatic cancers[78]. In dogs, the most common form is analogous to human non-Hodgkin lymphoma, specifically diffuse large B-cell lymphoma (DLBCL)[20,31]. Human non-Hodgkin lymphoma is a major health concern, with an incidence of 20/100,000 in the US[74]. An increased prevalence of different

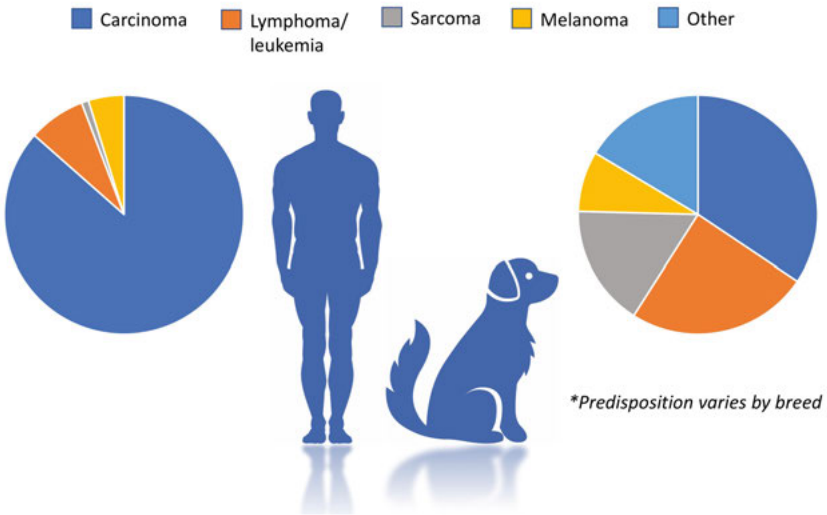

Figure 1. Different frequencies of cancer types between humans and dogs. Numbers adapted from NCI Surveillance, Epidemiology, and End Results (SEER) Report[74] and Schiffman and Breen, 2015[77]. Note that the frequency of certain cancer types will vary by breed, and that sarcomas are actually even rarer in humans (about 1% of cancers) than depicted here.

immunophenotypes of lymphoma has been noted in certain dog breeds, indicating that heritable risk factors for these subtypes are likely enriched in these breeds[79]. Lymphoma is often highly responsive to chemotherapy (commonly using a CHOP protocol) with 80-90% of dogs achieving complete remission, however, dogs eventually relapse and succumb to the disease[20,31]. Perturbation of common pathways between human and canine DLBCLs have been found; for example, the NF-KB pathway[80], and the PI3K pathway[81].

Osteosarcoma (OSA) is a primary bone tumor[31]. This cancer is very similar in canine and human patients, with tumors from the two species being histologically indistinguishable[31,63,66]. In both species, the tumors most commonly present in the appendicular skeleton, with metastasis to the lungs early in the clinical course[4,31,63,66]. OSA is much more common in large breed dogs than in humans[4]. Age of onset in the dog is bimodal, with peaks at 18-24 months and 7-9 years, while it is primarily a pediatric tumor in humans (median age 16 years)[66]. In both species, chemotherapy with platinum agents has been found to prolong survival after amputation or limb sparing surgery[4,82]. Human and canine tumors also have been shown to have similar genetic abnormalities, including mutations of *TP53*, *PTEN*, and *RB*, among other genes[31].

Numerous other canine cancers have the potential to be excellent models of human cancers, including melanoma, histiocytic sarcoma, transitional cell carcinoma, and squamous cell carcinoma[19,31,63,68,70]. I will explore the comparative aspects of canine hemangiosarcoma and human angiosarcoma later in this summary.

Neurodegenerative diseases

Neurodegenerative diseases are a diverse set of disorders with a shared underlying pathology, the progressive loss of neuronal structure and or function. These disorders include amyotrophic lateral sclerosis (ALS), Alzheimer’s disease, Parkinson’s disease, Huntington’s disease, and spinocerebellar ataxias, among others[83]. Many neurodegenerative diseases have both a Mendelian form as well as complex polygenic forms, in addition to environmental risk factors over the patient’s lifetime[83]. Most of these diseases are genetically heterogeneous within the population[83]. The pathophysiology of many neurodegenerative diseases include abnormal accumulation of proteins in neurons - for example, beta-amyloid in Alzheimer’s disease, α-synuclein in Lewy bodies in Parkinson’s disease, and SOD1 or TDP-43 aggregates in ALS[83,84]. For many of these diseases, palliative care is the only option[83] and new therapies are desperately needed for these patients.

ALS is a neurodegenerative disease characterized by degeneration specifically of the motor neurons, leading to weakness, muscle atrophy, and ultimately paralysis, with paralysis of the respiratory muscles being fatal[85–87]. Onset generally occurs late in life, with an average age of onset of approximately 60 years[88]. Patients can present with bulbar or spinal symptoms, and can have predominantly upper or lower motor neuron involvement (or both)[83,87]. In addition, some ALS patients experience cognitive impairment, as the spectrum of symptoms of ALS and fronto-temporal dementia (FTD) overlap[83,87]. Once onset symptoms occur, they are always progressive, although different patients progress at different rates[83–87].

Neuropathology in ALS patients report muscle atrophy, loss of motor neurons in the anterior horn, sclerosis and axon loss in the lateral corticospinal (descending motor) tracts in the spinal cord, and gliosis[84,89]. Although ALS is increasing in prevalence with our aging population, and is intensively studied, the mechanism behind the neuronal death and loss of nerve function is not yet known[89]. Possible contributing mechanisms include: oxidative stress, neuroinflammation, impairment of protein clearance, glutamate (excitatory) toxicity, abnormal RNA processing, defects in axonal and endosomal transport, defects in DNA-repair, mitochondrial dysfunction, and loss of oligodendrocytes[89].

ALS etiology is even more poorly understood. Environmental factors are thought to interact with genetic predisposition and play an important role in ALS etiology, particularly given that spatial clusters of cases have been found[88]. However, proving a causal relationship between environmental and lifestyle factors and ALS has been difficult, with many studies not showing a statistically significant association, or contradicting one another[89]. Some of the environmental factors that have been linked to a possible increased risk include: cigarette smoking, occupational exposure to heavy metals, electric shock, traumatic brain injury, air pollution, pesticides, and vigorous exercise[88].

Canine model of neurodegenerative disease

Pet dogs suffer from many inherited neurodegenerative diseases that have the potential to be models for human disease. Canine cognitive dysfunction, common in older dogs, has many similarities to human Alzheimer’s disease, including the presence of beta-amyloid plaques[90–92]. Numerous breeds suffer from hereditary ataxias, and spinocerebellar ataxia in particular, including Jack Russell Terriers, smooth-haired fox terriers, and Brittany spaniels[93]. Neuronal ceroid lipofuscinosis, a lysosomal storage disease, occurs in many breeds, and some of the causative mutations have been found in the same genes as in human patients[94,95]. Dogs also develop degenerative myelopathy, a disease with similar clinical presentation and progression as ALS in humans[96].

Pet dogs have the potential to be an important model for neurodegenerative diseases. Similar to the strengths of the dog as a model for human cancers, the strengths of the dog as a model for neurodegenerative disease include the decreased genetic heterogeneity within breeds, as well as, their shared environment with humans, which may help to strengthen some of the suggested associations to environmental factors in human studies. In addition, dogs may provide a unique insight into the progressive pathology of these diseases, as owners often elect euthanasia[97]. This may allow for evaluation of affected tissues from earlier in the disease progression than would normally be available from human patients[97].

Genetic mapping of complex traits in the dog

Over the past decade the domestic dog (*Canis familiaris*) has emerged as an important model organism for comparative and translational research in many complex diseases, including cancer[2–4,17–19,53,63,64,98,99].

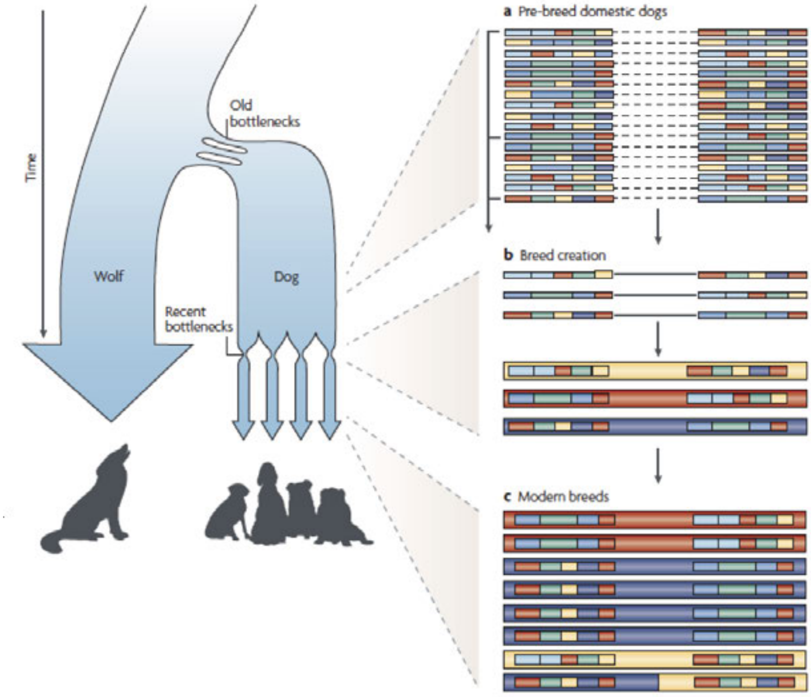

Figure *2.* Two population bottlenecks, one at domestication from the wolf, one more recently during creation of breeds, have led to a unique population structure in the domesticated dog. (a) Ancestral canine chromosomes. (b) During modern breed creation a smaller subset of the overall dog genetic diversity was selected in the founders of each breed. (c) Since breed creation was relatively recent, haplotypes within breeds are relatively long. Older ancestral haplotypes have had longer to recombine, thus are shorter. *From Karlsson & Lindblad-Toh, Nat Rev Genet, 2008[2]. Reprinted with permission from Springer Nature.*

The evolutionary history of the dog has led to a unique population structure, which can be used to aid disease gene mapping through genome-wide association studies (GWAS)[2,3,100]. As a species, dogs have undergone two major population bottlenecks – the first during domestication from the wolf at least 15,000 years ago, and the second during breed creation much more recently (in the last two hundred years, Figure 2)[2,3].

This natural history has led to dogs having long linkage disequilibrium (LD) and haplotype blocks within breeds, and short LD and haplotype blocks across breeds[2,3,99,101]. The LD within breeds is much longer than that found in human populations, and the haplotype diversity is lower, so the entire dog genome can be studied with far fewer SNP markers[2,3,99,102]. In addition, because breeds are closed, somewhat inbred populations, genetic diseases tend to be associated with particular breeds[15,99]. Because of this increased disease prevalence and lower genetic diversity within breeds, it is believed that disease alleles with higher risk might be more common in dogs than in humans, requiring fewer individuals to map such disease alleles[2,103]. This approach has been successful in identifying risk loci in complex, polygenic disorders in dogs, such as systemic lupus erythematosus (SLE)[103], atopic dermatitis[104], hemangiosarcoma and lymphoma[105], and osteosarcoma[72].

Methods

Genome-wide association studies (GWAS)

With the publication of the human genome[106] and the development of next-generation sequencing technologies, new complex disease studies and associations have increased exponentially, from genome-wide association studies to efforts directed at comprehensive mapping of the genetic and epigenetic landscape of each tumor type[107–109]. These efforts have led to rapid improvement in the diagnosis, molecular classification, and treatment of many cancers through the discovery of novel driver mutations, pathways, and biomarkers[42,107,108]. As available tools have allowed for more complex investigation of the cancer genome, the underlying complexity of cancer genomics and epigenomics has also begun to become clearer, necessitating ongoing development of new techniques and analysis methods[107,108].

One powerful tool for discovering germline risk loci across the genome has been genome-wide association studies (GWAS)[7,10,13]. In this method, thousands to millions of single nucleotide polymorphisms (SNPs) distributed across the entire genome are genotyped[10]. In a typical binary approach, the genotypes of patients with a given condition being studied ("cases") are compared to the genotypes of a control population. Association with quantitative phenotypes is also possible. Statistical tests (usually a Pearson's chi-square test) of genotype distribution between cases and controls are performed in order to locate SNPs with an association to the trait in question[110–112].

GWAS evolved from earlier genetic linkage studies involving study of the segregation of markers near a candidate gene within a family with a heritable disease[112]. This approach was successful in mapping many Mendelian diseases, but less successful in mapping common diseases where many genetic risk factors contribute a small amount each to overall disease risk[112,113]. In rare diseases, rare variants with high penetrance (or effect size) cause disease. However, the genetic architecture of common diseases is different, in that, while there are rare variants of high effect size, many common variants with a small effect size contribute to disease risk[110,112–114]. This is called the "common disease common variant hypothesis[110,114]."

GWAS has several strengths in the study of common, complex diseases, including the ability to interrogate variation across the entire genome with no prior knowledge of the genes or pathways involved in a particular disease[112,113], and coverage of much of the common variation in a population[110,112]. GWAS SNPs are chosen to be relatively common and polymorphic across the population of interest, and to "tag" much of the common variation across the genome by being in linkage disequilibrium with other variants not included in the GWAS. Linkage disequilibrium, as described briefly above, is a measure of how strongly correlated the alleles at two SNPs are[111]. The LD between two SNPs decreases over time due to genetic recombination. It is typically measured using either the population genetics measure D' or the statistical measure r^2 [110,111]. Therefore, the associated SNP in a GWAS may not be the causal variant, but may be tagging the causal variant through LD and further sequencing and functional studies may be required to determine the causal variant[110,112].

While GWAS has many strengths, there are also some limitations to this type of study. First, not all variants in the genome will be tagged by the selected SNPs. In particular, rare variants will not be tagged by the common SNPs selected for GWAS[102]. Because numerous SNPs are being tested, the probability of false positive associations is increased and must therefore be corrected for to determine at what level an association can be considered significant genome-wide[110]. There are numerous possible p-value corrections to determine genome-wide significance, including the conservative Bonferroni correction, false discovery rate (FDR), and permutation testing[110]. In addition, data must be carefully quality controlled to be sure that errors in genotyping, batch effects, or non-random missingness do not create spurious associations[111]. Population structure between cases and controls can also cause false associations, due to genetic differences between the two groups that are not associated with disease status[10,115]. This is important in dog breeds, where cryptic relatedness may be present due to breeding practices, and subpopulations may exist within the breed, such as "show" lines and "working" lines. For this reason, it is important to select cases and controls with similar genetic backgrounds from the same population. Stratification can be tested and corrected for by performing principle components analysis (PCA), and using a mixed linear model to perform the association[110,116].

Over the past decade or so of genome-wide association studies, it has become increasingly clear that larger sample sizes are required to detect association of numerous small-effect size loci in complex diseases[7,13]. One method that has been successfully used is the meta-analysis, whereby the summary statistics of multiple GWAS can be combined into a larger study. Meta-analysis methods for case-control studies combine the summary statistics (such as odds ratio (OR)) of each SNP from each study and report a combined effect[117,118]. Inter-study heterogeneity should be tested for, and can be controlled for by using a fixed-effects model (if differences are due to sampling error) or mixed- or random-effects model (if differences between the two populations exist)[117–120]. Genomic control is another method of correcting for inter-study heterogeneity or population structure, in which a fixed amount of inflation is corrected for across all SNPs, but this may not be appropriate unless all possible causes of population structure have been investigated[119,120].

Whole-genome imputation

Genotype imputation is a method for inferring the genotype of an individual at an untyped variant, based on the individual's genotype at typed variants, and a panel of many densely-genotyped reference individuals (reviewed in [121]). This approach is often used in genome-wide association studies in order to increase the number of SNPs that can be tested for association to the trait of interest while not incurring the cost of whole-genome sequencing of the entire cohort[121]. There are many methods currently in use, including IMPUTE2[122–124], BEAGLE[125], fastPHASE[126], and MACH[127]. These tools use different approaches to identify short stretches of haplotypes which are shared between individuals via identity by descent (IBD), and "fill in the gaps" of less densely genotyped samples by modeling them as patchworks of the haplotypes present in the population[121]. IMPUTE2, which we used in our HSA genome-wide meta-analysis, uses a method based on Hidden Markov Models (HMM). In human studies, these methods have been shown to perform well using large reference panels of haplotypes, such as those from the HapMap[128] or 1000 Genomes projects[121,129].

Whole-genome imputation is not yet widely used in canine GWAS, however, a recent paper by Friedenberg and Meurs demonstrated the method's feasibility in this species[130]. Indeed, the long stretches of linkage disequilibrium and low heterogeneity within dog breeds makes the method feasible with smaller reference panels, with a mixed-breed panel containing ~20 individuals of the breed of interest was found to perform well for imputation within a single breed, while a mixed-breed reference panel containing only 5 individuals from the breed of interest did not perform as well[130].

Pathway analysis

Pathway analysis has become a ubiquitous tool in gleaning biological meaning from many types of genomic studies. It can be applied to many different questions and types of data. In this thesis, I have used pathway analysis to look for functional enrichment in the risk loci found in our GWAS, to examine differences in gene expression by genotype in RNA-seq data, and to look for common pathways in the significantly mutated genes in our tumor/normal exome sequencing study.

In looking at the GWAS risk loci, all genes within an LD window of the top associated SNP at a given locus were input. Because the risk variants contribute only a small to moderate amount to overall risk, we must consider how these risk loci are interacting with each other to produce the disease phenotype[131–133]. Testing for pathway association can also lessens the burden of multiple-testing, as combining SNPs into pathways reduces the number of tests which must be performed[134,135]. However, this approach does not allow for genes to be ranked or prioritized.

In our initial GWAS of HSA/BLSA risk, we looked at how carrying the risk allele at the top associated GWAS loci affected gene expression in tumor samples. By calculating which genes were differentially expressed using the EdgeR[136,137] program, we were able to then use the fold-change in expression and p-values in the pathway analysis, which was performed using Ingenuity Pathway Analysis[138].

In our exome sequencing study of HSA, we used the DAVID[139,140] and STRING Database[141] tools to look for common pathways between our significantly mutated genes. However, this analysis was hampered by the low number of significantly mutated genes and the strong association of the top genes to many similar cancer pathways.

Many different tools exist to perform pathway (or gene set) analysis. Many of these tools use a "knowledge base" or public database of pathways (such as KEGG[142], GO[143], and MSigDB[144]), then compare user-identified gene lists to these pathways to test for enrichment (reviewed in[135,145]). Different approaches include over-representation analysis, functional class scoring, and pathway topology approaches[135], which Khatri, *et al.* classify as first, second, and third generation approaches, respectively[145]. Over-representation methods are perhaps the most straightforward, simply counting the number of genes in the input list that appear in a given pathway, and testing for enrichment[135,145]. However, this approach omits information in that it does not consider significance information, such as fold-change in expression or p-value[135,145]. The functional class scoring or gene-set approach improves on these methods by considering and scoring all genes in the experiment before calculating a pathway-wide score[135,145]. Newer topology-based methods have also emerged, which take into account the interactions of members of each pathway[135,145].

While pathway analysis methods have a number of advantages, as outlined above, it is important to note that there are also limitations to this type of analysis. In particular, the analysis is only as good as the underlying annotation, and despite a large amount of detailed functional data, we still do not have information on cell-type, the regulatory effects of non-coding variants, and the changes in pathways that occur under different environmental perturbations or disease states[145]. In addition, analysis will necessarily be biased toward the larger, more well-studied pathways and diseases (*e.g.* cancer)[134], while some genes within those pathways may have pleiotropic effects.

Whole exome sequencing

Whole exome sequencing is a type of targeted sequencing that has allowed for unbiased sequencing studies of the coding parts of the genome. By targeting the coding exons, studies can look for variants of high effect (missense or nonsense mutations) in patients with a particular disease, without *a priori* knowledge of what specific genes may be involved, and without the cost associated with sequencing the whole genome[146]. Several hybridization techniques (some solid-phase, some liquid-phase) are available, all employing a similar approach[146]. Briefly, biotinylated probes complementary to the genomic regions of interest are synthesized, and fragmented genomic DNA is exposed to these probes and hybridize with complementary sequences. The hybridized sequences are captured using streptavidin magnetic beads, and the DNA of interest eluted[146,147]. This approach has proved fruitful in identifying causal mutations for many disorders, both Mendelian and complex[146–148]. In addition, whole exome sequencing plays an important role in clinical evaluation of cancers, allowing identification of somatic driver mutations that could potentially be treated with targeted molecular therapies. However, it is important to note that this approach only covers approximately 1% of the genome, and will miss the majority of non-coding mutations, which have also been shown to play an important role as somatic driver mutations[149]. For example, mutations in the promoter of the *TERT* gene, which encodes telomerase reverse transcriptase, have been found to play a role in melanomas[150,151].

Sequencing of somatic variants

The application of precision medicine will require the routine calling and characterization of somatic variants from patient tumor sequencing data. However, somatic variants present unique challenges to traditional variant-calling techniques.

Germline variant callers expect a diploid sample with allelic fractions of 100% in the case of homozygotes, or 50% in the case of heterozygotes. However, the allelic fractions in sequencing data from tumor tissue are not nearly so clear-cut. Tumor samples can contain contamination with normal stromal tissue, which adds normal germline DNA to the somatic DNA from tumor cells[152,153]. In addition, because tumors grow by clonal expansion and evolution (through acquisition of new somatic mutations)[154], a given mutation may be found in only a subset of tumor cells[152,155]. Also, the sample cannot be assumed to be diploid, as many cancer genomes include amplifications or deletions[152,156]. Thus, while rare errors or artifacts can easily be ignored by germline callers, in somatic calling, they could easily be confused with real variants at low allelic fraction[152]. In order to call somatic variants with greater confidence, it is therefore necessary to sequence tumor samples to a greater depth than normal samples (typically 60-100x)[157]. In addition, many algorithms use paired (*i.e.* from the same individual) normal sequencing data in order to eliminate germline variants from the list of somatic variant calls[152]. However, it is important to note that there can be some level of tumor DNA in the normal samples - for example, due to circulating tumor cells and cell-free DNA in a whole blood sample, or due to collection of normal tissue adjacent to tumor tissue[158].

Before somatic variants can be called, it is important to properly prepare the data in order to minimize artifacts. The Genome Analysis Toolkit (GATK) best practices[159] recommend marking PCR duplicates using the Picard[160] tool MarkDuplicates, and aligning the reads to the reference genome using BWA[161]. Finally, base quality scores should be recalibrated using a set of known variants[159]. Once these preprocessing steps are done, variant calling can proceed.

Many different somatic mutation-calling programs have been written using different strategies to identify true somatic variants and filter out false positives, and focusing on different types of variants. MuTect2, which was used for the tumor/normal exome sequencing study reported in this thesis, uses a haplotype-based approach to somatic variant calling which involves local realignment around insertions/deletions, construction of possible haplotypes, and examining read support for each possible haplotype[152]. It also does not assume diploidy in the data, evaluating the likelihood that a given variant was caused by artifact or is a true variant, and producing a log-likelihood score (LOD) of the variant being truly somatic[152,162]. Variants found in a panel of normals created by calling variants on all matched normal samples in the study are filtered out, along with variants at known germline variant locations. MuTect2 identifies both SNVs and indels.

There are numerous sources of error and artifact in next-generation sequencing of tumor samples, including PCR bias, which can decrease coverage of GC and AT-rich regions of the genome[163]; polymerase errors, when a mismatched nucleotide is incorporated during PCR[164]; and sequence-specific sequencing errors (SSE) in the Illumina platform which can cause false SNV calls in certain contexts[165].

In order to limit the number of false positive variant calls, filters are applied to the final variant set. Costello, *et al.*, reported artifactual G>T transversions in NGS sequencing data due to oxidation of guanine during acoustic shearing of the DNA in library preparation[166]. These artifacts can be minimized by transferring DNA to Tris-EDTA buffer prior to shearing[166]. In addition, many of these artifacts can be removed by applying a filter based on strand bias and the LOD of the variant[166].

Formalin fixation has been shown to cause several sequencing artifacts[167]. The DNA obtained from formalin-fixed paraffin embedded (FFPE) archival tissue samples is highly fragmented, reducing PCR efficiency[167]. In addition, formalin causes deamination of cytosine, leading to artifactual C>T transversions[167,168], which can be reduced during library preparation by treating DNA with uracil-DNA glycosylase[169]. Low pH can also result in abasic sites after purine (adenine and guanine) nucleotides are cleaved from the backbone, leading to a decrease in usable DNA as well as artifactual SNVs and deletions[167,170]. The GATK tool FilterMutectCalls has recently implemented an experimental filter to try to reduce the number of artifactual C>T transversions in data from FFPE samples.

In 2016, the National Cancer Institute (NCI) reported on its Genomic Data Commons website that large numbers of artifactual insertions had been called in samples from the TCGA in which whole genome amplification had been performed and in which variant calling was performed using MuTect2[171]. These insertions were supported only by "soft-clipped" reads (reads in which some number of bases at one end had not aligned along with the rest of the read, and were therefore "clipped" by the aligner), and had no other support[171]. Rerunning MuTect2 using the "--dontUseSoftClippedBases" option resolved the artifact[171]. We saw a similar artifact in our tumor/normal exome sequencing study, in which large numbers of insertions were called supported only by soft clipped reads, almost exclusively in FFPE samples. This was resolved using the "--dontUseSoftClippedBases" option in MuTect2. Although we are unsure of the exact mechanism, the WGA insertion artifacts were thought to be caused by chimeric reads formed during the WGA process[172]. We hypothesize that formalin cross-linking may have led to chimeric reads in our FFPE samples.

Functional evaluation of genetic variants

Most of the variants found to be associated with disease in GWAS have been non-coding, and are presumed to have a regulatory effect on gene expression[10,173]. However, our understanding of the effect of non-coding variants on expression is still quite limited. Large-scale efforts such as the Encyclopedia of DNA Elements (ENCODE) project have provided annotation of many functional elements in the human genome, including DNase I hypersensitivity sites, epigenetic marks, and transcription factor binding sites in different cell lines[174]. In addition, the 29 Mammals Project used a comparative genomics approach to identify regions of the genome that are under evolutionary constraint[175]. These resources can serve to prioritize candidate functional variants, and to generate hypotheses as to their function. However, determining the actual role each variant plays is a painstaking process, requiring detailed functional analysis.

RNA-seq

RNA-seq is a method for sequencing all or a subset of the RNA transcripts in a given cell or tissue[176]. It is commonly used as a method of sequencing all messenger RNA (mRNA) in a given tissue sample to perform an analysis of the expression levels of the various transcripts. Briefly, the method involves isolating the RNA of interest (often by selecting for the poly-adenylated 3' tail in the case of mRNAs[177]), converting the RNA to cDNA, and then proceeding with library construction and next-generation sequencing[176]. Transcriptome analysis using RNA-seq can play an important role in elucidating the effects of a given variant on gene expression, which may lead to better understanding of how the variant affects disease susceptibility[178,179]. Indeed, there are now large-scale efforts underway to identify expression quantitative trait loci (eQTLs) across the human genome in the Genotype-Tissue Expression (GTEx) project[180]. Data from this ongoing project has shown that 92% of common GWAS variants were associated with gene expression changes[181]. RNA-seq can also be used to identify germline SNPs or somatic variants within the coding regions of genes[182,183], and can also be used to investigate whether a given variant has an effect on the splicing of the RNA and thus the end protein product[184]. RNA-seq is also an important method for finding gene fusions in cancers. Although RNA-seq is very powerful, it does have some limitations. RNA-seq data derived from tissue samples is a mixture of many different cell types, which may reduce power to detect changes in expression or variants that occur only in particular cell types[185]. This problem can be addressed using newer single-cell sequencing methods, however, the tissue of interest must be known *a priori* in this case. Coverage of different genes in RNA-seq data will be variable and subject to dynamic processes[182]. For example, it would be difficult to validate a mutation in RNA-seq data if the mutation induces nonsense-mediated decay[186] of the transcript of interest. In addition, RNA is very unstable, and samples must be handled carefully in order to prevent RNA degradation[178].

Electrophoretic Mobility Shift Assay (EMSA)

EMSA is a tool for investigating the interaction of proteins (transcription factors, regulatory complexes, etc) with DNA sequences[187–190]. One potential use of this technique is to investigate whether differences in binding occur due to a sequence variant[187]. Briefly, oligonucleotides containing the wild-type and mutant sequence can be synthesized, and incubated with nuclear extract from the cell type of interest. Gel electrophoresis is performed using an agarose or polyacrylamide gel[189], and the subsequent gel dyed and imaged[187,189]. Protein-DNA complexes will be larger than either the protein or DNA individually, and thus migrate more slowly through the gel[190,191], allowing identification of differential binding. EMSA is a simple but powerful technique, however, it does not identify which proteins are binding. If a particular protein is suspected to be binding, a supershift assay can be performed using antibodies to that protein - the antibodies will bind the protein-DNA complex, further slowing migration through the gel[190,192]. One consideration with this technique is the choice of nuclear extract, as the protein of interest may be expressed in a cell-type or treatment dependent manner.

Clinical variables

Statistical analyses

Analyses of the effect of a certain exposure or treatment on survival or other time-to-events of interest (diagnosis of disease, discharge from hospital, etc) require special consideration. If the distribution of events over time is not normal, nonparametric tests should be used, and the median used as a measure of central tendency[193,194].

Kaplan-Meier survival curves are a useful tool in studies of clinical outcome, as they allow time-to-event to be calculated even when the event of interest has not been observed in some study participants[194–196]. In participants where the event of interest has not been observed, the time during which they were under observation is still of interest, so these patients can still be considered in the analysis, and are noted as "right censored," meaning that the event of interest had not occurred by the end of the study period[195,196]. The probability of a participant having the event of interest in a given time period is calculated based on the number of participants at risk, and the times at which events occurred[194–196]. The differences between two survival curves can be calculated using the log-rank test[194,196]. The Cox proportional hazards model can be used to investigate the effect of a numeric or categorical variable on the time-to-event, and to estimate the risk of the event per unit of time[194,196].

Consideration of survival time bias

Another important consideration when comparing retrospective treatment outcomes is the concept of "immortal time" bias. This bias occurs when patients are divided into treatment groups retrospectively, and inclusion in a given group has a time component[197,198]. For example, it is not recommended to start chemotherapy until approximately two weeks after surgery to allow time for the surgical incision to heal. Therefore, any patient retrospectively included in the surgery and chemotherapy group must have survived those two intervening weeks. This type of bias must be corrected for in order to produce a more accurate estimate of the effect of a treatment or exposure[197,198].

Canine hemangiosarcoma

Biology of HSA and comparative aspects with angiosarcoma (AS)

Hemangiosarcoma (HSA), is an aggressive cancer of the vascular endothelium with a rapid and aggressive clinical course[31]. Blood-filled tumors can form anywhere in the vasculature with canine patients often showing no clinical signs until the tumor ruptures, leading to severe internal hemorrhage and sudden death[199,200]. Because of the "silent" nature of this cancer and its aggressive metastasis, most patients have advanced disease by the time of diagnosis[199,200]. Standard of care includes surgical resection of the tumor and adjuvant chemotherapy (usually doxorubicin-based)[17]. However, a recent retrospective study of outcomes in splenic hemangiosarcomas showed that, while adjuvant chemotherapy improves survival over surgery alone in the early follow-up period (4 months), it may not prolong survival overall[201].

Canine HSA could be a potentially important model for human angiosarcoma. Angiosarcoma is rare in humans, but is similarly difficult to treat in humans as hemangiosarcoma is in canines, and carries a poor prognosis[202–205]. The rarity of angiosarcoma in humans precludes many genetic studies. Canine HSA is naturally occurring and histologically and clinically very similar to human angiosarcoma[203], and the fact that it is so common makes it a good candidate for study.

The etiology of this cancer in dogs is mostly unknown, but, as in human angiosarcoma, both environmental factors and hormonal factors have been linked to the disease. Beagles that were exposed to inhaled radiation showed a high incidence of hepatic, pulmonary, and bone HSA[206]. In addition, it is theorized that ultraviolet radiation exposure plays a role in cutaneous HSA, as it tends to occur in light pigmented dogs and in areas with little fur[207]. Spaying has been associated with an increased risk of cardiac[208] and splenic hemangiosarcoma[209], and neutering with an increased risk of cardiac hemangiosarcoma[208]. In humans, exposure to vinyl chloride, thorium dioxide, arsenicals, and androgens are known risk factors, however, the cancer occurs much less frequently in humans[76,210–212].

Genetic studies of human AS and canine HSA

Recent genomic, immunohistochemistry, and other studies have shown that HSA may arise from hematopoietic precursors in the bone marrow expressing hematopoietic markers such as c-KIT and CD34 and endothelial markers such as CD146[202,213]. Genomic studies of canine HSA have shown increased expression in tumors of growth factors VEGF and *FGF2*, *VEGF* receptors flt-1 and flk-1, RB1 and its pathway member cyclin D1[214], the angiogenic protein Angiopoeitin-2[215], the anti-apoptotic protein survivin[216], pro-inflammatory cytokine IL-8[217], signal transducer STAT3[218], preproendothelin-1 (PPET-1) and endothelin type A receptor (ETA)[219]. Downregulation/deletion of the genes encoding tumor suppressors CDKN2A/p16[214,220] and PTEN[221] have also been documented. HSA tumors were found not to express COX-2[222]. *Ras* and *VHL* mutations were found to be rare[223]. Endothelin receptor antagonists were found to inhibit HSA cell growth *in vitro*[219]. One study of human AS found that the PIK3CA/AKT/mTOR pathway was activated, despite not finding *PTEN* mutations[224]. This pathway is also active in canine HSA[225].

A survey by Thomas, *et al.*, of somatic copy number variations (SCNAs) in five breeds predisposed to HSA showed extensive heterogeneity, including limited aneuploidy and copy number variation which was not highly shared across cases. The most common SCNAs were loss of chromosome 16, loss of chromosome 11 near the *CDKN2A/B* locus, and gains on chromosomes 13, 24 and 31. Gain of the *MYC* oncogene was the only recurrent SCNA shared between studies of canine and human angiosarcoma. In this study, the tumors from golden retrievers were more similar to each other than to other breeds, suggesting an interplay between the genetic background and cytogenetic tumor profile[220].

The rarity of AS in humans has precluded large cohort studies, however, there have been several small exome or targeted sequencing studies. Behjati, *et al.* performed a combination of WGS (n=3), WES (n=8) and two targeted sequencing panels (n=4, n=24)[226]. They found recurrent mutations in the genes phospholipase C gamma 1 (*PLCG1,* 3/34 cases) and protein tyrosine phosphatase, receptor type B (*PTPRB,* 10/39 cases), both of which play roles in angiogenesis[226]. Whole exome sequencing of two family members with Li-Fraumeni-like syndrome who developed cardiac angiosarcoma revealed a shared variant in protection of telomeres 1 (*POT1*)[227]. A targeted sequencing study of 34 AS found frequent *TP53* mutations, as well *PTPRB*, *PLCG1*, and mutations in the MAPK pathway[228]. Interestingly, another study of 62 angiosarcomas found only 2/52 cases had *TP53* mutations, while 13/33 cases had a frameshift mutation in *PIK3CA*[224]. However, visualization of numerous studies using the COSMIC Cancer Browser reveals that *TP53* mutations were the most common mutation in AS (28/106. 26%), while *PIK3CA* mutations were rare (2/67, 3%)[229]. This may be due to differences in the population being studied, tumor location, or environmental exposure.

A recent WES study of 20 HSA cases from various breeds of dog found *PIK3CA* mutations in 9/20 cases, *TP53* mutations in 7/20 cases, and *PTEN* mutations in 2/20 cases[230]. Eight of the nine PIK3CA mutations were found in amino acid 1047, a known *PIK3CA* mutation hotspot in human cancers, with seven having the H1047R mutation commonly seen in human cancers[230]. In addition, they identified one *PLCG1* mutation with a mutation homologous to a S345F mutation previously reported in human AS[230].

Canine degenerative myelopathy

Biology of DM and comparative perspective with ALS

Numerous animal models have been developed to study the various forms of ALS including mouse, rat, zebrafish, fly, nematode, yeast, and pig[84,89]. However, the dog is the only known mammal in which a comparable disease occurs spontaneously[97]. Canine degenerative myelopathy is a progressive neurodegenerative disease leading to gradual loss of motor function and many other clinical similarities to upper motor neuron onset forms of ALS in humans[96]. Dogs typically present with hindlimb proprioceptive deficits and weakness, progressing to paresis and paralysis, and ultimately tetraplegia and failure of the respiratory muscles[96]. Neuropathology changes include: axon degeneration of the descending motor and ascending sensory tracts, demyelination of the spinal cord white matter, astrocyte proliferation[96], and intracytoplasmic neuronal inclusion bodies similar to those seen in human ALS.

Genetic studies of human ALS and canine DM

In addition to being clinically heterogeneous, ALS is highly genetically heterogeneous in the population. While mutations in over 30 genes have been linked to the disease[89,231], within a given patient, the disease architecture is thought to be monogenic or oligogenic[83,232]. Both familial and sporadic forms of ALS are seen, with the majority (90-95%) of cases being sporadic[83,87,233]. The first mutation linked to ALS was in the superoxide dismutase 1 (*SOD1*) gene[234], which was found to be causal in approximately 20% of familial cases with a mutation in a known ALS gene[233], but is rare in sporadic cases. Patients with *SOD1* mutations have neuronal inclusions consisting of SOD1 protein aggregates, while most other ALS patients have inclusion bodies made up of TDP-43 aggregates. A hexanucleotide expansion in the *C9ORF72*[235,236] gene causes approximately 40% of familial cases and 7% of sporadic cases[88]. Other causal genes include *TARDBP*[237] (encoding TDP-43), *FUS*[238,239], *OPTN*[240], and *UBQLN2*[241].

Our group, along with collaborators, performed a GWAS of DM cases in Pembroke Welsh corgis, with finemapping in four other predisposed breeds (boxer, German Shepherd Dog, Chesapeake Bay retriever, and Rhodesian ridgebacks), identifying a strong association to a *SOD1* missense mutation, in a homologous position to a known human ALS mutation[97]. A second *SOD1* mutation was discovered shortly afterwards, only carried by Bernese mountain dogs[242]. To date, no other genes have been causally linked to canine DM. Although DM was found to be autosomally recessively inherited (in contrast to ALS, which is most commonly autosomal dominant), mutant canine SOD1 was found to form active dimers which aggregate, supporting a toxic gain-of-function as in human *SOD1*-mutant ALS[243]. However, the *SOD1* mutation was incompletely penetrant, suggesting that either genetic or environmental modifiers play a role in disease pathogenesis[97].

Aims of the thesis

The overall goal of the project was to characterize the genetics of multiple canine complex traits. Specifically:

Hemangiosarcoma

- Characterize the germline risk factors contributing to the high risk of hemangiosarcoma in the golden retriever breed.
- Examine clinical and genetic data for associations of genotype with clinical presentation or outcome.
- Characterize the tumor somatic mutational landscape of canine hemangiosarcoma from a comparative oncology perspective.
- Look for correlation between germline, clinical, and somatic features.

Degenerative myelopathy

- Characterize genetic modifiers of risk for degenerative myelopathy in *SOD1* mutant homozygous Pembroke Welsh corgis.
- Perform functional analysis of variants at the top risk loci to elucidate the role of *SP110* variants in disease pathophysiology.

Summary of included papers

Part 1 - Germline risk factors in hemangiosarcoma

Background

In the first two papers, our goal was to map germline risk factors predisposing golden retrievers to the high rates of hemangiosarcoma in the breed. We collected histologically confirmed (or strongly suspected based on clinical findings) cases of hemangiosarcoma in golden retrievers from samples submitted directly to the Broad Institute by owners and veterinarians, or samples collected by our collaborators at University of Minnesota or North Carolina State University. In addition, we collected samples from healthy controls that were 10 years or older and were cancer free. In Paper II, we expanded the study to include additional cases and controls, and investigated the link between genetic risk and clinical presentation and outcome in these patients.

Paper I: Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers

Collected cases and controls were genotyped using the Illumina Canine HD Whole-Genome Genotyping BeadChip, with over 170,000 markers across the canine genome. After filtering SNPs for genotyping and missingness, a GWAS was performed using a linear mixed model in GCTA. We discovered a strong association to HSA on chromosome 5, as well as less significant associations on chromosomes 11, 13, 22, 25, and 33. Further analysis of the chromosome 5 locus revealed two separate associations located approximately 4 Mb apart, one at 29 Mb and another at 33 Mb, which together explain approximately 20% of disease risk. Interestingly, the risk loci on chromosome 5 were found to predispose dogs to both HSA and BLSA. Each locus was found to contain two separate risk haplotypes. The two risk haplotypes at 29 Mb were mostly inherited together, and did not segregate enough to determine whether they tag more than one causative variant. The association seen in HSA cases became stronger when BLSA cases were added. The haplotypes at 29 Mb were relatively common, with nearly 50% of all cases and 25% of controls homozygous for the risk haplotype. At the 33 Mb locus, we found one shared haplotype associated with both HSA and BLSA, and a second haplotype associated only with BLSA predisposition. The second risk haplotype was only seen together with the shared 33 Mb risk haplotype and was much longer, indicating it happened later and on the shared 33 Mb risk haplotype. Both haplotypes at 33 Mb were rare compared to the 29 Mb locus – only 4% of HSA cases and none of the controls were homozygous for the shared risk haplotype.

Finemapping of the region was performed in nine individuals, including cases and controls. No coding variants were found in the associated regions which were inherited with the shared risk haplotypes. RNA-seq of 22 BLSA tumors revealed that carrying the risk haplotype at the 29 Mb region was associated with a strong *cis*-regulatory effect on the expression of the gene *TRPC6*. This gene encodes a transient receptor cation channel involved in calcium regulation, which has a role in many processes, including T-cell activation[244,245]. The shared risk haplotype at the 33 Mb region was associated with a *trans*-regulatory effect on many genes. Ingenuity Pathway Analysis of the differentially expressed genes associated with the 33 Mb shared risk haplotype suggested that the top associated pathways are important for T-cell differentiation, activation, and signaling. The mechanism of how these immune regulatory risk loci predispose golden retrievers to HSA and BLSA is under investigation. We theorize that it may be related to tumor immune surveillance.

Paper II: Genome-wide meta-analysis identifies inherited variation contributing to overall risk and age of onset in canine angiosarcoma

The aim of this study was to explore the relationship between genetic risk and clinical characteristics of HSA in golden retrievers (GR), with a particular focus on genetic or clinical factors associated with age of onset or survival after diagnosis. We performed a second GWAS of HSA in the GR breed, using an additional 64 HSA cases and 96 cancer-free controls, and performed a meta-analysis of the two HSA GWAS. Genome-wide significant associations were found on seven chromosomes, with the top SNP on chromosome 18 ($p = 1.62 \times 10^{-6}$). The second most associated region was located on chromosome 5 ($p = 3.84 \times 10^{-6}$), confirming our earlier GWAS finding. In addition, less significant peaks confirmed the previously described chromosome 3 and chromosome 11 associations, as well as illuminating new associations on chromosomes 24, 15, and 2.

The top SNP on chromosome 5 in the meta-analysis was the same as the top SNP in the original cohort ($p_{original} = 3.85 \times 10^{-6}$, $p_{meta} = 3.84 \times 10^{-6}$). The top SNP at the previously described 33 Mb locus was less strongly associated in the meta-analysis ($p_{original} = 4.02 \times 10^{-6}$, $p_{meta} = 6.44 \times 10^{-4}$). The meta-analysis also confirmed the association on chromosome 11 at 37.7 Mb, and the association on chromosome 3 at 83.4 Mb that was seen in the combined HSA-BLSA analysis.

The most significant association in the meta-analysis was a new risk locus on chromosome 18, which had not been seen in the original GWAS. Several interesting candidate genes lie within and near the associated region, including the genes encoding histone methyltransferase KMT5B, which has been shown to play a role in telomere recombination[246,247], the demethylase KDM2A, the cell cycle checkpoint protein RAD9A, and the CDK2 inhibitor CDK2AP2.

Medical records were examined, and clinical variables noted. These included signalment, age of diagnosis of HSA, survival time, tumor location, packed red-cell volume (PCV), mitotic rate, and treatment protocol. The referring veterinarian or owner was contacted when possible to complete missing information. We then looked at the association of clinical variables with each other, as well as with the top GWAS risk loci. The risk genotype at the top SNP on chromosome 18 was associated with earlier age of onset ($p = 0.0094$).

In addition, the total number of risk alleles from each of the top associated SNPs had a more significant association with age of onset, with dogs with 0 risk alleles ($n = 5$) getting the disease at a median age of 9.66 years, and dogs with 9 risk alleles ($n = 3$) getting the disease at a median age of 6.96 years ($p = 0.0009$). When the number of risk alleles was dichotomized into Low and High groups, and weighted by OR, the association with age of onset became more significant ($p = 0.0003$).

The location of the tumor was most strongly associated with survival time ($p = 1.85 \times 10^{-9}$), with the cutaneous/subcutaneous group surviving dramatically longer than all other groups ($\text{median}_{sq} = 393$ days, $n = 14$). The type of treatment a dog received also made a significant difference in survival time, with both the surgery only group ($p = 6.3 \times 10^{-4}$) and the surgery with adjuvant chemotherapy group ($p = 1.3 \times 10^{-4}$) surviving significantly longer than the group that received no treatment.

Discussion

In these two studies, we identified 7 genetic loci associated with hemangiosarcoma risk in golden retrievers, and found that two shared risk loci on chromosome 5 also predispose these dogs to B-cell lymphoma. The chromosome 5 risk loci were also associated with downregulation of *TRPC6* in *cis*, and 100 genes throughout the genome in *trans*, which were enriched for pathways associated with T-cell number and activation. This suggests a possible mechanism by which these risk factors predispose goldens to cancer - they may suppress T-cell immunosurveillance and anti-tumor activity. In addition, we were able to investigate the effect of genotype at the GWAS risk loci on clinical presentation and outcome. The fact that dogs with a greater genetic risk burden are diagnosed earlier would suggest that these risk factors interact, at least in part, in an additive manner. These studies highlight the advantages of leveraging the unique genetic structure of pure-bred dogs to investigate the mechanisms by which genetic risk factors influence the clinical presentation and outcome of a rare human cancer.

Part 2 – Somatic mutations in hemangiosarcoma

Background

Having mapped germline risk factors for hemangiosarcoma, our next goal was to characterize the somatic mutational landscape of this tumor. To do this, we performed a whole exome sequencing study of tumor and paired normal germline DNA from whole blood collected from 47 golden retrievers with visceral hemangiosarcoma. The tumor samples were either frozen (n = 17), or formalin-fixed and paraffin-embedded (FFPE, n = 30), and were either splenic, right atrial, or liver tumors.

Paper III - Exome sequencing of hemangiosarcomas in the golden retriever reveals frequent mutation of the *TP53* tumor suppressor and *PIK3CA* oncogene

In this study, we set out to understand HSA from a different angle by examining the somatic variants that arise in HSA tumors. We performed a whole exome sequencing study of tumor and paired normal DNA in 47 golden retrievers. We performed exome capture using the Roche Nimblegen system and sequenced all samples to a target depth of 60x in the tumor and 30x in the normal on the Illumina HiSeq. The average sequencing depth was 78x in the tumor and 63x in the normal. After preprocessing following the GATK best practices, variants were called using MuTect2[162]. To obtain a more conservative set of variants, we took the consensus of the somatic calls from the GATK3 and GATK4 versions of MuTect2. Coding mutations were functionally annotated using SNPeff[248], and Genome MuSiC[249] was used to identify significantly mutated genes.

Seven genes were significantly mutated in our cohort of golden retrievers (Table 1). These included tumor suppressor *TP53* (28/47 cases, 59.5%), oncogene *PIK3CA* (14/47 cases, 29.8%) and its regulatory subunit *PIK3R1* (4/47 cases, 8.5%), Origin recognition complex subunit 1 (*ORC1*), which plays a role in DNA replication (4/47 cases, 8.5%), RAS p21 protein activator 1 (*RASA1*, 4/47 cases, 8.5%), actin related protein 2/3 complex subunit 1A (*ARPC1A,* 3/47 cases, 6.3%), and ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d (ATP5H, 2/47 cases, 4.25%). Ten of the fourteen *PIK3CA* mutations occur at amino acid position 1047, a mutational hotspot frequently mutated in many types of human cancers.

Gene	**Indels**	**SNVs**	**Tot Muts**	**Covd Bps**	**Muts pMbp**	**P-value**	**FDR**
TP53	3	25	28	72582	385.8	0	0
PIK3CA	0	14	14	156415	89.5	0	0
PIK3R1	0	4	4	124857	32.0	7.29×10^{-8}	5.88×10^{-4}
ORC1	0	4	4	133219	30.0	1.35×10^{-7}	8.18×10^{-4}
RASA1	0	4	4	136539	29.3	3.38×10^{-6}	1.64×10^{-2}
ARPC1A	0	3	3	72351	41.5	1.26×10^{-5}	5.08×10^{-2}
ATP5H	0	2	2	22074	90.6	1.71×10^{-5}	5.90×10^{-2}

Table 1. Significantly mutated genes in golden retriever angiosarcoma tumors.

Twenty-three cases (48.9%) had at least one mutation in a gene in the PI3K pathway (including catalytic subunits *PIK3CA*, *PIK3CB*, *PIK3C2G*, and *PIK3C3*, regulatory subunits *PIK3R1* and *PIK3R5*; and *PTEN*). These mutations tended not to co-occur. We also saw amplifications of the catalytic subunits and deletions of the regulatory subunits and *PTEN*. Heart tumors had the most frequent *PIK3CA* mutations (11/16, 68.8%), with only 9/22 (40.9%) of splenic cases having this mutation. *PIK3CA* mutations were much rarer in liver tumors (1/8, 12.5%) and did not reach significance in that tumor subgroup alone. However, the overall number of PI3K mutations in liver tumors was 37.5%, and this cohort had a very small sample of liver tumors.

In humans, recurrent mutations in phospholipase C gamma 1 (*PLCG1*) have been reported[226]. While we did not see recurrent mutations in this gene (we saw one mutation in this gene in our cohort), we did note that 8 cases (17%)

have mutations in other phospholipase C genes, which included *PLCB2, PLCB4, PLCD4, PLCE1, and PLCG2*. Recurrent protein tyrosine phosphatase receptor type B (*PTPRB*) mutations have also been reported in human angiosarcomas[226]. While we did not see mutations in this gene in our cohort, we did note mutations in other protein tyrosine phosphatase genes in seven (14.9%) cases. These mutations in related genes may be causing a similar phenotype; indeed, *PTPRD* and *PTPRS*, which are mutated in our cohort, have previously been reported mutated in a study of human angiosarcomas[228].

In addition, we examined the association of germline and somatic mutations, which may be easier to identify in the limited genetic diversity within a single dog breed. We found that cases carrying the rare GWAS risk allele at chromosome 11 (near the gene SH3 domain containing GRB2 like 2, endophilin A1(*SH3GL2*)), were less likely to have a mutation in the PI3K pathway, although our sample size was not large enough for this to be statistically significant. Two somatic mutations were seen in the paired box protein 3 (*PAX3*) gene, which is close to the GWAS chromosome 2 risk locus, suggesting that this gene may be affected both at a germline and somatic level.

Discussion

In this study, we harnessed the power of the limited genetic diversity within a single dog breed to investigate the landscape of somatic mutations in canine HSA. It is interesting to note that, while we found similarities between human AS and canine HSA including *TP53* mutations, we also see differences, such as the large number of *PIK3CA* mutations in the canine tumors, when this gene has only rarely been reported as mutated in human AS. Since we know that the PI3K pathway is active in this cancer in both species, this suggests that the PI3K pathway tends to be activated at different points in the canine and human tumors, with the same downstream result. The phosphatidylinositol-3-kinase (PI3K) pathway plays an important role in signal transduction for cellular proliferation, differentiation, and survival, and is commonly mutated in many human cancers[250,251]. It is also interesting to note that while *PLCG1* and *PTPRB* are not recurrently mutated in our cohort, we see recurrent mutations in related genes which may lead to the same phenotype. Since angiosarcoma is known to be a heterogenous cancer in humans. Using the limited genetic diversity within dog breeds, it is possible that we may be able to characterize different molecular subtypes, and begin to connect germline and somatic mutations with clinical presentation and outcome.

Part 3 - Modification of risk in canine degenerative myelopathy with *SOD1* mutation

Background

While the last paper in this thesis is not cancer-focused, it offered a unique opportunity to explore the relationship between genetic variation and clinical presentation in another complex disease. In this study, we investigated the variable penetrance and age of onset of degenerative myelopathy (DM) in Pembroke Welsh corgis (PWC) with a *SOD1* mutation that had previously been linked to DM. DM is a neurodegenerative disorder characterized by gradual loss of motor neurons and eventual paralysis, which has many clinical and pathological similarities to ALS in humans. We discovered variants within the *SP110* nuclear body protein gene that increase risk of developing DM and decrease age at presentation.

Paper IV - Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy

In a previous GWAS study of canine degenerative myelopathy, we and collaborators discovered a *SOD1* mutation that increased the risk of developing degenerative myelopathy in several dog breeds, including the PWC[97]. However, some dogs who were homozygous for the risk mutation developed disease early in life, while others developed it later, or never developed disease at all. To investigate, we performed a GWAS of 15 cases and 31 controls, all of which were PWC and homozygous for the *SOD1* risk allele. The GWAS was performed using a mixed model in EMMAX[252], and revealed a clear association on chromosome 25, with the top associated SNP having a p-value of $P = 2.7 \times 10^{-8}$.

We performed finemapping by whole-genome sequencing 3 PWCs (2 cases, 1 control). This allowed us to add 101 new SNPs from the nearby region for genotyping in the original and replication GWAS cohorts. The new genotyping was merged with the original data, and missing data was imputed using BEAGLE[125]. This analysis revealed 5 significantly associated SNPs. Haplotype phasing revealed that these 5 associated SNPs formed 4 haplotypes, and the risk haplotype was carried by the majority (9/15, 60%) of cases, and was found in only one control.

In the original GWAS cohort, which was selected to have extreme phenotypes - younger cases and older controls, the frequency of the risk haplotype was significantly different between cases and controls ($p = 1.7 \times 10^{-5}$). We attempted to replicate the association in a new cohort of 32 cases and 13 control PWC, but this did not reach statistical significance because the age at onset was not as different between the two groups as in the original cohort. However, the difference in the frequency of the risk haplotype between cases and controls increased in significance ($p = 1.5x10^{-5}$) when we merged the original and replication cohorts, replicating the association.

We performed a Kaplan–Meier survival analysis of PWCs with and without the risk haplotype, measuring time to DM diagnosis in the combined cohort of all PWC DM cases and controls. Individuals who did not develop signs were censored at the last time point when information was available.

The risk haplotype is within the gene *SP110* on chromosome 25. *SP110* encodes the SP110 nuclear body protein. There were two coding mutations within the haplotype region, which were predicted to be silent. Two additional variants were interesting functional candidates because they overlapped a hotspot for transcription factor binding. We investigated the functional effects of these variants using electrophoretic mobility shift assays (EMSAs), luciferase assays, and RNA-seq, and found that variants altered transcription factor binding, reduced *SP110* expression in a human T-cell line, and were associated with alternative splicing of *SP110*.

Discussion

Our GWAS in the Pembroke Welsh Corgi breed revealed variants within the *SP110* gene that modify the risk of developing DM and age at diagnosis in dogs homozygous for the *SOD1* risk mutation. *SP110* is strongly expressed in immune cells, and is a member of the SP100/SP140 family of nuclear body proteins[253,254]. This protein plays a role in many important cellular functions from transcription, to DNA damage response, to infection, and even apoptosis[255]. Our findings suggest that the immune response may influence disease onset in DM, and highlight the need for further investigation of what role SP110 and the DNA-sensing pathway plays in the development of DM in other breeds, as well as in ALS in humans.

General discussion

We have shown in these four studies that dogs are an excellent model for investigating the underlying biology of complex diseases, including both cancer and neurodegenerative disorders. The limited diversity within dog breeds facilitates the identification of multiple interacting genetic risk factors, while the high rates of disease within breeds allows genetic risk factors to be more easily connected to variation in disease presentation and clinical outcome. Using this approach, we have discovered new genetic risk loci for hemangiosarcoma and degenerative myelopathy, which will hopefully bring new insight into the pathophysiology of these diseases, and may lead to new diagnostics and treatment strategies in both human and veterinary patients.

Investigating the genetic basis of complex disease

Together, our analyses have revealed several regions of the genome associated with risk of HSA in the golden retriever breed and shown that overall genetic risk has an impact on age of onset of the disease. In addition, we have elucidated one possible mechanism through which the two risk-associated loci on chromosome 5 predispose golden retrievers to HSA. In addition to playing a role in T-cell activation, TRPC6 is necessary for extravasation of leukocytes out of the blood vessels into surrounding tissue during the inflammatory response[256], suggesting that dogs carrying the risk allele on chromosome 5 at 29 Mb may have a defect in the inflammatory response. It would be interesting to further investigate whether this risk factor predicts response to immunotherapy, and whether it has an effect on other phenotypes, such as autoimmune disease in the golden retriever breed. In light of the results from the tumor/normal exome sequencing study confirming the importance of the PI3K pathway in golden retriever HSA, it is also interesting to note that the 29 Mb risk allele is also associated with a down-regulation (though not as dramatic as seen in *TRPC6*) of the gene encoding PI3K regulatory subunit *PIK3R6*, the regulatory subunit for *PIK3CG*. *PIK3R6* may be a tumor suppressor; knockdown of this gene led to increased Akt activation and lung metastases in a mouse carcinoma model[257]. Fully understanding the perturbations of the PI3K pathway in golden retrievers may lead to new strategies for targeted therapies, as monotherapy with non-isoform-specific PIK3CA inhibitors has proven disappointing for the treatment of human patients in the clinic[258].

Investigation of the interaction between germline variants and the somatic mutations that arise in the tumor may be more tractable within the limited genetic background of a dog breed. For example, we see a possible trend for dogs carrying the risk allele at the chromosome 11 GWAS locus not to also have somatic mutations in the PI3K pathway, suggesting that the risk allele at this locus (near the gene SH3 domain containing GRB2 like 2, endophilin A1(*SH3GL2*)), may already perturb this pathway sufficiently to mitigate the selective advantage of a somatic mutation in the same pathway. Using the publicly available human cancer data on CBioPortal[259], we searched mutations and copy number aberrations. We found that while somatic *SH3GL2* alterations are fairly rare, they are less likely to occur with *PIK3CA* mutations (OR=0.542, p<0.001). *SH3GL2* is a tumor suppressor, which plays a role in EGF receptor endocytosis[260] and its downregulation, activates the PI3K pathway[261]. We also see two somatic mutations in the *PAX3* gene in our tumor/normal study, which is very near the common chromosome 2 GWAS risk locus, and this may be an example of a gene contributing to disease risk via both germline and somatic mutations.

Of further interest in this thesis is the role of variants which may modify the immune response in the pathogenesis of both HSA and DM, neither of which are immune-mediated diseases. This highlights the importance of the immune system in shaping the environment in which complex diseases arise, and the layers of interacting risk factors that must be analyzed in order to fully understand disease risk. The dog is a particularly strong model system in which to discover these effects, as the canine immune system is very similar to the human immune system and these diseases are arising as polygenic traits on a simplified genetic background (within a breed).

The dog as a model of human complex diseases

By studying hemangiosarcoma in the golden retriever, we confirmed important similarities between the canine and human cancer, including *TP53* mutations and the importance of involvement of the PI3K pathway. However, we also see differences between the two species, such as the high frequency of *PIK3CA* mutations in goldens, which have not been commonly reported in human cases. It seems likely that canine HSA is still a good model for human AS, despite this difference. We know that the PI3K pathway is active in this cancer in both species, therefore it is likely that in humans, the PI3K pathway is altered at a different point - perhaps by *PTEN* deletion or mutation in a different catalytic subunit. It is also possible that HSA in the golden retriever is a good model only for a subset of human AS - the visceral sites included in our study are much rarer in humans, and may be underrepresented in the current small studies, which likely have higher numbers of the more common head and neck or breast angiosarcomas.

Another area of interest is the identification of canine cancers which are models for human cancers at the genomic level, even if they are not necessarily the same cancer or tissue type. One example of this is the fact that dogs commonly have a *BRAF* mutation in transitional cell carcinomas (TCC) homologous to the human V600E mutation[68,262]. In humans, the *BRAF* V600E mutation is common in malignant melanomas, and less common in other cancers[68,263]. Despite the fact that transitional cell carcinoma and melanoma are very different cancers histologically, canine TCC may be a good model more broadly for *BRAF* V600E mutant cancer[68,262]. Looking at our hemangiosarcoma exome data from this perspective, there are similarities to human serous endometrial adenocarcinoma, which also shows frequent *TP53* and *PIK3CA* mutation, with occasional *PIK3R1* mutations, and without *PTEN* mutation[264]. Whether canine HSA would be a good model of serous endometrial adenocarcinoma, another highly vascular cancer, is worth exploring further.

Our investigation into the incomplete penetrance of the *SOD1* mutation in canine degenerative myelopathy led us to discover an interesting modifier gene, which may provide insight into the complex and poorly-understood pathophysiology of both DM and ALS. Neuroinflammation is one of the mechanisms theorized to play a role in these diseases. It is well established that microglia are activated in ALS patients, and that there are elevations in proinflammatory cytokines[265]. However, it is not known whether this inflammation is a cause of the disease, or a secondary response to degenerating neurons[260]. Studies in *SOD1*-mutant mice have shown that blocking production of the mutant protein in microglial cells slows progression of the disease, and co-culturing motor neurons with *SOD1*-mutant microglia decreased survival of the neurons[265]. Our finding that *SP110*, which plays a role in the innate immune system's DNA-sensing pathway, modifies DM risk suggests that variation in the immune response may indeed modify the course of DM. It would be therapeutically advantageous to investigate this pathway further in human *SOD1*-mutant ALS patients. We may only be able to see this clear modifier effect because of the limited genetic diversity within the PWC breed, and it may be that PWC DM is a particularly good model for a subset of human ALS cases.

Future directions

We are continuing to explore the data from these studies in order to gain a more complete understanding of the pathophysiology of hemangiosarcoma in golden retrievers.

In the genome-wide meta-analysis, we have performed whole-genome imputation of the HSA GWAS data in the program IMPUTE2 using a reference panel of 353 dogs (including 13 golden retrievers). In total, 8,394,795 variants were imputed with high confidence (info metric $\geqq$ 0.7, individual uncertainty < 0.1). We are continuing to process this data, and will have a genome-wide imputed GWAS soon. We will then confirm the accuracy of the top imputed variants by performing genotyping in a subset of GWAS samples via Sanger sequencing.

Once we have candidate variants fine-mapped through imputation, we will proceed with functional analyses to identify the causal variant(s) at each locus and the corresponding functional effect. We plan to begin this analysis by mapping expression quantitative trait loci (eQTLs) within each associated region, comparing differential expression between patients carrying the risk allele at each locus and those carrying the non-risk allele. This analysis will be performed only within the LD window of the associated region to look for *cis* regulatory effects, which will hopefully mitigate some of the noise inherent in whole-tissue RNA-seq data from such a heterogeneous tumor type. In the future, we would like to perform single-cell RNA-seq in hemangiosarcomas in order to more clearly define the expression profiles of the tumor cells.

It is our hope that in the coming years a large, prospective study can be undertaken to investigate the predictive value of these risk loci in golden retrievers, and whether a useful test could be designed in order to identify dogs at high risk for the disease who might benefit from additional screening, or even prophylactic measures. In addition, we will continue to investigate the relationship between the germline risk factors and the somatic mutational profiles of the tumors, to see whether germline risk factors are predictive of what genes or pathways are affected somatically, which will be of great value both in terms of understanding the pathophysiology of the disease, as well as to potentially guide treatment decisions.

We are now investigating liquid biopsy - or sequencing tumor genomes from circulating cell-free tumor DNA in the blood of patients - as a method of collecting tumor samples. We have already performed a pilot study in 27 dogs showing that this approach is also feasible in dogs, and well-suited to a highly vascular tumor like hemangiosarcoma. By sequencing tumor genomes directly from blood samples, we will be able to dramatically increase our sample numbers, as surgical biopsy samples can be difficult to obtain, and not every patient goes to surgery. In addition, liquid biopsy also opens up the possibility of routine longitudinal sampling to guide therapeutic decisions by monitoring minimal residual disease and the emergence of chemotherapy-resistant clones.

Acknowledgments

I would like to take this opportunity to express my gratitude to the many people at Uppsala University, the Broad Institute, and beyond without whom this work would not have been possible.

Kerstin, thank you for being a wonderful, brilliant, and patient advisor. You first introduced me to genomics, and I've never wanted to do anything else since. Thank you for sharing your vision with me. I've been honored to be able to study with you.

Ginger, you are one of the kindest (and smartest) souls I have ever met. Thank you for all your help and advice these past 4 years. I will miss working with you, but I hope you will visit as often as you can! I can't wait to meet the new little one!

Elinor, you are the best "advisor in spirit" ever! Thank you for taking me under your wing, for all your help and support, and for being the brilliant scientist you are. I'm looking forward to many exciting studies ahead!

Hyun Ji, thank you for being an amazing friend and officemate! You always had time to consult about science or listen to me worry. Our office is going to seem very empty without you.

Diane, thank you for also being an amazing officemate - for your scientific input, your editing skills, and for many trips to get tea.

Ross, thank you for being brilliant and patient, and for invaluable help in the wetlab.

Jason, thank you for being brilliant and patient, and for invaluable help with various bioinformatics tools.

Jeremy, thanks for not actually hiding under your desk when I came to ask for help getting samples sequenced, or for a cost object, or any number of other random things.

Michele, thank you for all your help with sample wrangling, DNA extraction, PCR, health updates, and a million other things.

Uppsala cancer group **-** **Maja**, **Vikki**, **Sharda**, **Iris**, and **Chao**, thank you for many excellent discussions and journal clubs.

Mom and **Dad**, thank you for being the most loving and supportive parents anyone could ever ask for! I think after 25+ years in school, I am finally done getting degrees.

Tresi, thank you for being so loving, patient, and supportive. I would probably have starved if you weren't such an amazing cook.

Zoe, you are the best cat. Thank you for helping me type by snoozing on the trackpad, and for being adorable.

References

1. Schork, N. J. Genetics of complex disease: approaches, problems, and solutions. *Am. J. Respir. Crit. Care Med.* **156,** S103–9 (1997).
2. Karlsson, E. K. & Lindblad-Toh, K. Leader of the pack: gene mapping in dogs and other model organisms. *Nat. Rev. Genet.* **9,** 713–725 (2008).
3. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–819 (2005).
4. Withrow, S. J. & Wilkins, R. M. Cross talk from pets to people: translational osteosarcoma treatments. *ILAR J.* **51,** 208–213 (2010).
5. LaRue, S. M. *et al.* Limb-sparing treatment for osteosarcoma in dogs. *J. Am. Vet. Med. Assoc.* **195,** 1734–1744 (1989).
6. Fu, W., O'Connor, T. D. & Akey, J. M. Genetic architecture of quantitative traits and complex diseases. *Curr. Opin. Genet. Dev.* **23,** 678–683 (2013).
7. Price, A. L., Spencer, C. C. A. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc. Biol. Sci.* **282,** 20151684 (2015).
8. Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265,** 2037–2048 (1994).
9. Marian, A. J. Molecular genetic studies of complex phenotypes. *Transl. Res.* **159,** 64–79 (2012).
10. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322,** 881–888 (2008).
11. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19,** 110–124 (2018).
12. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45,** D896–D901 (2017).
13. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101,** 5–22 (2017).
14. Parker, H. G. & Ostrander, E. A. Canine genomics and genetics: running with the pack. *PLoS Genet.* **1,** e58 (2005).
15. Patterson, D. F. Companion animal medicine in the age of medical genetics. *J. Vet. Intern. Med.* **14,** 1–9 (2000).
16. Association, A. V. M. & Others. US pet ownership & demographics sourcebook. in (American Veterinary Medical Association, 2012).
17. Khanna, C. *et al.* The dog as a cancer model. *Nat. Biotechnol.* **24,** 1065–1066 (2006).

18. Khanna, C., London, C., Vail, D., Mazcko, C. & Hirschfeld, S. Guiding the optimal translation of new cancer treatments from canine to human cancer patients. *Clin. Cancer Res.* **15,** 5671–5677 (2009).
19. Rowell, J. L., McCarthy, D. O. & Alvarez, C. E. Dog models of naturally occurring cancer. *Trends Mol. Med.* **17,** 380–388 (2011).
20. Marconato, L., Gelain, M. E. & Comazzi, S. The dog as a possible animal model for human non-Hodgkin lymphoma: a review. *Hematol. Oncol.* **31,** 1–9 (2013).
21. Kelsey, J. L., Moore, A. S. & Glickman, L. T. Epidemiologic studies of risk factors for cancer in pet dogs. *Epidemiol. Rev.* **20,** 204–217 (1998).
22. Reif, J. S. Animal sentinels for environmental and public health. *Public Health Rep.* **126 Suppl 1,** 50–57 (2011).
23. Reif, J. S., Bruns, C. & Lower, K. S. Cancer of the nasal cavity and paranasal sinuses and exposure to environmental tobacco smoke in pet dogs. *Am. J. Epidemiol.* **147,** 488–492 (1998).
24. Reif, J. S., Dunn, K., Ogilvie, G. K. & Harris, C. K. Passive smoking and canine lung cancer risk. *Am. J. Epidemiol.* **135,** 234–239 (1992).
25. Glickman, L. T., Domanski, L. M., Maguire, T. G., Dubielzig, R. R. & Churg, A. Mesothelioma in pet dogs associated with exposure of their owners to asbestos. *Environ. Res.* **32,** 305–313 (1983).
26. Hayes, H. M., Jr, Hoover, R. & Tarone, R. E. Bladder cancer in pet dogs: a sentinel for environmental cancer? *Am. J. Epidemiol.* **114,** 229–233 (1981).
27. Hayes, H. M. *et al.* Case-control study of canine malignant lymphoma: positive association with dog owner's use of 2,4-dichlorophenoxyacetic acid herbicides. *J. Natl. Cancer Inst.* **83,** 1226–1231 (1991).
28. Hayes, H. M., Tarone, R. E., Casey, H. W. & Huxsoll, D. L. Excess of seminomas observed in Vietnam service U.S. military working dogs. *J. Natl. Cancer Inst.* **82,** 1042–1046 (1990).
29. Tarone, R. E. *et al.* Service in Vietnam and risk of testicular cancer. *J. Natl. Cancer Inst.* **83,** 1497–1499 (1991).
30. Weinberg, R. *The Biology of Cancer, Second Edition*. (Garland Science, 2013).
31. Withrow, S. J. & Page, R. L. *Withrow and MacEwen's Small Animal Clinical Oncology*. (Elsevier Health Sciences, 2013).
32. Robert, J. Comparative study of tumorigenesis and tumor immunity in invertebrates and nonmammalian vertebrates. *Dev. Comp. Immunol.* **34,** 915–925 (2010).
33. Patel, L. R., Nykter, M., Chen, K. & Zhang, W. Cancer genome sequencing: understanding malignancy as a disease of the genome, its conformation, and its evolution. *Cancer Lett.* **340,** 152–160 (2013).
34. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).
35. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

36. Evans, S. C. & Lozano, G. The Li-Fraumeni syndrome: an inherited susceptibility to cancer. *Mol. Med. Today* **3,** 390–395 (1997).
37. Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 820–823 (1971).
38. Heisey, R. E., Carroll, J. C., Warner, E., McCready, D. R. & Goel, V. Hereditary breast cancer. Identifying and managing BRCA1 and BRCA2 carriers. *Can. Fam. Physician* **45,** 114–124 (1999).
39. Levy-Lahad, E. & Friedman, E. Cancer risks among BRCA1 and BRCA2 mutation carriers. *Br. J. Cancer* **96,** 11–15 (2007).
40. Venkitaraman, A. R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108,** 171–182 (2002).
41. Roberts, N. J. & Klein, A. P. Genome-wide sequencing to identify the cause of hereditary cancer syndromes: with examples from familial pancreatic cancer. *Cancer Lett.* **340,** 227–233 (2013).
42. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
43. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
44. Lindahl, T. & Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11,** 3610–3618 (1972).
45. Seebode, C., Lehmann, J. & Emmert, S. Photocarcinogenesis and Skin Cancer Prevention Strategies. *Anticancer Res.* **36,** 1371–1378 (2016).
46. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21,** 7435–7451 (2002).
47. Lutz, W. K. Endogenous genotoxic agents and processes as a basis of spontaneous carcinogenesis. *Mutat. Res.* **238,** 287–295 (1990).
48. Tomlinson, I., Sasieni, P. & Bodmer, W. How many mutations in a cancer? *Am. J. Pathol.* **160,** 755–758 (2002).
49. Oncogenes and tumor suppressor genes | American Cancer Society. Available at: https://www.cancer.org/cancer/cancer-causes/genetics/genes-and-cancer/oncogenes-tumor-suppressor-genes.html. (Accessed: 25th March 2018)
50. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Metastasis: dissemination and growth of cancer cells in metastatic sites. *Nat. Rev. Cancer* **2,** 563 (2002).
51. Corrie, P. G. Cytotoxic chemotherapy: clinical aspects. *Medicine* **36,** 24–28 (2008).
52. Brown, C. Targeted therapy: An elusive cancer target. *Nature* **537,** S106–8 (2016).
53. Ranieri, G. *et al.* A model of study for human cancer: Spontaneous occurring tumors in dogs. Biological features and translation for new anticancer therapies. *Crit. Rev. Oncol. Hematol.* **88,** 187–197 (2013).
54. Rangarajan, A. & Weinberg, R. A. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat. Rev. Cancer* **3,** 952–959 (2003).

55. Sharpless, N. E. & Depinho, R. A. The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat. Rev. Drug Discov.* **5,** 741–754 (2006).
56. Shultz, L. D. *et al.* Human cancer growth and therapy in immunodeficient mouse models. *Cold Spring Harb. Protoc.* **2014,** 694–708 (2014).
57. Walsh, N. C. *et al.* Humanized Mouse Models of Clinical Disease. *Annu. Rev. Pathol.: Mech. Dis.* **12,** 187–215 (2017).
58. Hidalgo, M. *et al.* A pilot clinical study of treatment guided by personalized tumorgrafts in patients with advanced cancer. *Mol. Cancer Ther.* **10,** 1311–1316 (2011).
59. Michell, A. R. Comparative clinical science: The medicine of the future. *Vet. J.* **170,** 153–162 (2005).
60. Gorbunova, V. *et al.* Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19392–19396 (2012).
61. Tian, X. *et al.* High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499,** 346–349 (2013).
62. Caulin, A. F. & Maley, C. C. Peto’s Paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* **26,** 175–182 (2011).
63. Vail, D. M. & MacEwen, E. G. Spontaneously occurring tumors of companion animals as models for human cancer. *Cancer Invest.* **18,** 781–792 (2000).
64. Paoloni, M. & Khanna, C. Translation of new cancer treatments from pet dogs to humans. *Nat. Rev. Cancer* **8,** 147–156 (2008).
65. Shearin, A. L. & Ostrander, E. A. Leading the way: canine models of genomics and disease. *Dis. Model. Mech.* **3,** 27–34 (2010).
66. Morello, E., Martano, M. & Buracco, P. Biology, diagnosis and treatment of canine appendicular osteosarcoma: similarities and differences with human osteosarcoma. *Vet. J.* **189,** 268–277 (2011).
67. Pinho, S. S., Carvalho, S., Cabral, J., Reis, C. A. & Gärtner, F. Canine tumors: a spontaneous animal model of human carcinogenesis. *Transl. Res.* **159,** 165–172 (2012).
68. Gardner, H. L., Fenger, J. M. & London, C. A. Dogs as a Model for Cancer. *Annu Rev Anim Biosci* **4,** 199–222 (2016).
69. Glickman, L., Glickman, N., Thorpe - Retriever Club of America National …, R. & 1999. The Golden Retriever Club of America National Health Survey 1998-1999. *grca.org* (1999).
70. Davis, B. W. & Ostrander, E. A. Domestic dogs and cancer research: a breed-based genomics approach. *ILAR J.* **55,** 59–68 (2014).
71. Shearin, A. L. *et al.* The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. *Cancer Epidemiol. Biomarkers Prev.* **21,** 1019–1027 (2012).
72. Karlsson, E. K. *et al.* Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol.* **14,** R132 (2013).

73. Song, R. B., Vite, C. H., Bradley, C. W. & Cross, J. R. Postmortem evaluation of 435 cases of intracranial neoplasia in dogs and relationship of neoplasm with breed, age, and body weight. *J. Vet. Intern. Med.* **27,** 1143–1152 (2013).
74. Lynn A.G. Ries, National Cancer Institute, Bethesda *et al.* SEER Cancer Statistics Review, 1975-2003. (2006).
75. Burningham, Z., Hashibe, M., Spector, L. & Schiffman, J. D. The epidemiology of sarcoma. *Clin. Sarcoma Res.* **2,** 14 (2012).
76. Penel, N., Marréaud, S., Robin, Y.-M. & Hohenberger, P. Angiosarcoma: state of the art and perspectives. *Crit. Rev. Oncol. Hematol.* **80,** 257–263 (2011).
77. Schiffman, J. D. & Breen, M. Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370,** (2015).
78. Priester, W. A. & McKay, F. W. The occurrence of tumors in domestic animals. *Natl. Cancer Inst. Monogr.* 1–210 (1980).
79. Modiano, J. F. *et al.* Distinct B-cell and T-cell lymphoproliferative disease prevalence among dog breeds indicates heritable risk. *Cancer Res.* **65,** 5654–5661 (2005).
80. Mudaliar, M. A. V. *et al.* Comparative gene expression profiling identifies common molecular signatures of NF-κB activation in canine and human diffuse large B cell lymphoma (DLBCL). *PLoS One* **8,** e72591 (2013).
81. Mooney, M. *et al.* Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of Canis familiaris. *PLoS One* **8,** e61088 (2013).
82. Phillips, B. *et al.* Use of single-agent carboplatin as adjuvant or neoadjuvant therapy in conjunction with amputation for appendicular osteosarcoma in dogs. *J. Am. Anim. Hosp. Assoc.* **45,** 33–38 (2009).
83. Pihlstrøm, L., Wiethoff, S. & Houlden, H. Chapter 22 - Genetics of neurodegenerative diseases: an overview. in *Handbook of Clinical Neurology* (eds. Kovacs, G. G. & Alafuzoff, I.) **145,** 309–323 (Elsevier, 2018).
84. Picher-Martel, V., Valdmanis, P. N., Gould, P. V., Julien, J.-P. & Dupré, N. From animal models to human disease: a genetic approach for personalized medicine in ALS. *Acta Neuropathol Commun* **4,** 70 (2016).
85. Al-Chalabi, A. & Leigh, P. N. Recent advances in amyotrophic lateral sclerosis. *Curr. Opin. Neurol.* **13,** 397–405 (2000).
86. Al-Chalabi, A. & Hardiman, O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat. Rev. Neurol.* **9,** 617–628 (2013).
87. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. *Lancet* **377,** 942–955 (2011).
88. Talbott, E. O., Malek, A. M. & Lacomis, D. Chapter 13 - The epidemiology of amyotrophic lateral sclerosis. in *Handbook of Clinical Neurology* (eds. Aminoff, M. J., Boller, F. & Swaab, D. F.) **138,** 225–238 (Elsevier, 2016).

89. Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nat Rev Dis Primers* **3,** 17085 (2017).
90. Romanucci, M. & Della Salda, L. Oxidative Stress and Protein Quality Control Systems in the Aged Canine Brain as a Model for Human Neurodegenerative Disorders. *Oxid. Med. Cell. Longev.* **2015,** 940131 (2015).
91. Giaccone, G. *et al.* Cerebral preamyloid deposits and congophilic angiopathy in aged dogs. *Neurosci. Lett.* **114,** 178–183 (1990).
92. Ishihara, T. *et al.* Immunohistochemical and immunoelectron microscopical characterization of cerebrovascular and senile plaque amyloid in aged dogs' brains. *Brain Res.* **548,** 196–205 (1991).
93. Urkasemsin, G. & Olby, N. J. Canine hereditary ataxia. *Vet. Clin. North Am. Small Anim. Pract.* **44,** 1075–1089 (2014).
94. Faller, K. M. E. *et al.* The Chihuahua dog: A new animal model for neuronal ceroid lipofuscinosis CLN7 disease? *J. Neurosci. Res.* **94,** 339–347 (2016).
95. Katz, M. L. *et al.* Enzyme replacement therapy attenuates disease progression in a canine model of late-infantile neuronal ceroid lipofuscinosis (CLN2 disease). *J. Neurosci. Res.* **92,** 1591–1598 (2014).
96. Nardone, R. *et al.* Canine degenerative myelopathy: a model of human amyotrophic lateral sclerosis. *Zoology* **119,** 64–73 (2016).
97. Awano, T. *et al.* Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 2794–2799 (2009).
98. Ostrander, E. A., Galibert, F. & Patterson, D. F. Canine genetics comes of age. *Trends Genet.* **16,** 117–124 (2000).
99. Sutter, N. B. & Ostrander, E. A. Dog star rising: the canine genetic system. *Nat. Rev. Genet.* **5,** 900–910 (2004).
100. Ostrander, E. A. & Wayne, R. K. The canine genome. *Genome Res.* **15,** 1706–1716 (2005).
101. Parker, H. G. *et al.* Genetic structure of the purebred domestic dog. *Science* **304,** 1160–1164 (2004).
102. Wang, W. Y. S., Barratt, B. J., Clayton, D. G. & Todd, J. A. Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.* **6,** 109–118 (2005).
103. Wilbe, M. *et al.* Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nat. Genet.* **42,** 250–254 (2010).
104. Tengvall, K. *et al.* Genome-wide analysis in German shepherd dogs reveals association of a locus on CFA 27 with atopic dermatitis. *PLoS Genet.* **9,** e1003475 (2013).
105. Tonomura, N. *et al.* Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers. *PLoS Genet.* **11,** e1004922 (2015).
106. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).

107. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153,** 17–37 (2013).

108. Wheeler, D. A. & Wang, L. From human genome to cancer genome: the first decade. *Genome Res.* **23,** 1054–1062 (2013).

109. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45,** 1113–1120 (2013).

110. Bush, W. S. & Moore, J. H. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* **8,** e1002822 (2012).

111. Lewis, C. M. & Knight, J. Introduction to genetic association studies. *Cold Spring Harb. Protoc.* **2012,** 297–306 (2012).

112. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

113. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6,** 95–108 (2005).

114. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17,** 502–510 (2001).

115. Knowler, W. C., Williams, R. C., Pettitt, D. J. & Steinberg, A. G. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am. J. Hum. Genet.* **43,** 520–526 (1988).

116. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46,** 100–106 (2014).

117. Nakaoka, H. & Inoue, I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *J. Hum. Genet.* **54,** 615–623 (2009).

118. Zeggini, E. & Ioannidis, J. P. A. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10,** 191–201 (2009).

119. Thompson, J. R., Attia, J. & Minelli, C. The meta-analysis of genome-wide association studies. *Brief. Bioinform.* **12,** 259–269 (2011).

120. Zeng, P. *et al.* Statistical analysis for genome-wide association study. *J. Biomed. Res.* **29,** 285–297 (2015).

121. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11,** 499–511 (2010).

122. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1,** 457–470 (2011).

123. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).

124. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44,** 955–959 (2012).

125. Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84,** 210–223 (2009).

126. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78,** 629–644 (2006).
127. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).
128. International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
129. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
130. Friedenberg, S. G. & Meurs, K. M. Genotype imputation in the domestic dog. *Mamm. Genome* (2016). doi:10.1007/s00335-016-9636-9
131. Hirschhorn, J. N. Genomewide association studies--illuminating biologic pathways. *N. Engl. J. Med.* **360,** 1699–1701 (2009).
132. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).
133. Mooney, M. A., Nigg, J. T., McWeeney, S. K. & Wilmot, B. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet.* **30,** 390–400 (2014).
134. Elbers, C. C. *et al.* Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet. Epidemiol.* **33,** 419–431 (2009).
135. Jin, L. *et al.* Pathway-based analysis tools for complex diseases: a review. *Genomics Proteomics Bioinformatics* **12,** 210–220 (2014).
136. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
137. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40,** 4288–4297 (2012).
138. Krämer, A., Green, J., Pollard, J., Jr & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30,** 523–530 (2014).
139. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).
140. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
141. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).
142. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

143. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
144. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).
145. Khatri, P., Sirota, M. & Butte, A. J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* **8,** e1002375 (2012).
146. Teer, J. K. & Mullikin, J. C. Exome sequencing: the sweet spot before whole genomes. *Hum. Mol. Genet.* **19,** R145–51 (2010).
147. Gasc, C., Peyretaillade, E. & Peyret, P. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and non-model organisms. *Nucleic Acids Res.* **44,** 4504–4518 (2016).
148. Shen, T., Pajaro-Van de Stadt, S. H., Yeat, N. C. & Lin, J. C.-H. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. *Front. Genet.* **6,** 215 (2015).
149. Gan, K. A., Carrasco Pro, S., Sewell, J. A. & Bass, J. I. F. Identification of Single Nucleotide Non-coding Driver Mutations in Cancer. *Front. Genet.* **9,** 16 (2018).
150. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339,** 957–959 (2013).
151. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339,** 959–961 (2013).
152. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* **16,** 15–24 (2018).
153. Krøigård, A. B., Thomassen, M., Lænkholm, A.-V., Kruse, T. A. & Larsen, M. J. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLoS One* **11,** e0151664 (2016).
154. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976).
155. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20,** 68–80 (2010).
156. Storchova, Z. & Pellman, D. From polyploidy to aneuploidy, genome instability and cancer. *Nat. Rev. Mol. Cell Biol.* **5,** 45–54 (2004).
157. Wilmott, J. S. *et al.* Tumour procurement, DNA extraction, coverage analysis and optimisation of mutation-detection algorithms for human melanoma genomes. *Pathology* **47,** 683–693 (2015).
158. Wei, L. *et al.* Pitfalls of improperly procured adjacent non-neoplastic tissue for somatic mutation analysis using next-generation sequencing. *BMC Med. Genomics* **9,** 64 (2016).
159. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).

160. Picard Tools - By Broad Institute. Available at: http://broadinstitute.github.io/picard. (Accessed: 23rd March 2018)
161. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
162. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–219 (2013).
163. Kozarewa, I. *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* **6,** 291–295 (2009).
164. Kebschull, J. M. & Zador, A. M. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.* **43,** e143 (2015).
165. Nakamura, K. *et al.* Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39,** e90 (2011).
166. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41,** e67 (2013).
167. Do, H. & Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin. Chem.* **61,** 64–71 (2015).
168. Williams, C. *et al.* A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am. J. Pathol.* **155,** 1467–1471 (1999).
169. Do, H. & Dobrovic, A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. *Oncotarget* **3,** 546–558 (2012).
170. Zsikla, V., Baumann, M. & Cathomas, G. Effect of buffered formalin on amplification of DNA from paraffin wax embedded small biopsies using real-time PCR. *J. Clin. Pathol.* **57,** 654–656 (2004).
171. MuTect2 Insertion Artifacts | NCI Genomic Data Commons. Available at: https://gdc.cancer.gov/content/mutect2-insertion-artifacts. (Accessed: 23rd March 2018)
172. Buckley, A. R. *et al.* Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC Genomics* **18,** 458 (2017).
173. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).
174. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306,** 636–640 (2004).
175. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).
176. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
177. Rio, D. C., Ares, M., Jr, Hannon, G. J. & Nilsen, T. W. Enrichment of poly(A)+ mRNA using immobilized oligo(dT). *Cold Spring Harb. Protoc.* **2010,** db.prot5454 (2010).

178. Cieślik, M. & Chinnaiyan, A. M. Cancer transcriptome profiling at the juncture of clinical translation. *Nat. Rev. Genet.* **19,** 93–109 (2018).

179. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16,** 197–212 (2015).

180. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).

181. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).

182. Goya, R. *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* **26,** 730–736 (2010).

183. Piskol, R., Ramaswami, G. & Li, J. B. Reliable identification of genomic variants from RNA-seq data. *Am. J. Hum. Genet.* **93,** 641–651 (2013).

184. Liu, J. *et al.* Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res.* **22,** 2315–2327 (2012).

185. Shannon, C. P., Yang, C. X. & Tebbutt, S. J. A Bloody Primer: Analysis of RNA-Seq from Tissue Admixtures. *Methods Mol. Biol.* **1712,** 175–201 (2018).

186. Ghosh, S. & Jacobson, A. RNA decay modulates gene expression and controls its fidelity. *Wiley Interdiscip. Rev. RNA* **1,** 351–361 (2010).

187. Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* **9,** 3047–3060 (1981).

188. Fried, M. G. & Liu, G. Molecular sequestration stabilizes CAP–DNA complexes during polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **22,** 5054–5059 (1994).

189. Lane, D., Prentki, P. & Chandler, M. Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol. Rev.* **56,** 509–528 (1992).

190. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.* **2,** 1849 (2007).

191. Cai, Y.-H. & Huang, H. Advances in the study of protein–DNA interaction. *Amino Acids* **43,** 1141–1146 (2012).

192. Kristie, T. M. & Roizman, B. Alpha 4, the major regulatory protein of herpes simplex virus type 1, is stably and specifically associated with promoter-regulatory domains of alpha genes and of selected other viral genes. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3218–3222 (1986).

193. Pagano, M. & Gauvreau, K. *Principles of Biostatistics, Second Edition.* (CRC Press, 2018).

194. Benítez-Parejo, N., Rodríguez del Águila, M. M. & Pérez-Vicente, S. Survival analysis and Cox regression. *Allergol. Immunopathol.* **39,** 362–373 (2011).

195. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **53,** 457–481 (1958).

196. George, B., Seals, S. & Aban, I. Survival analysis and regression models. *J. Nucl. Cardiol.* **21,** 686–694 (2014).
197. Suissa, S. Immortal time bias in pharmaco-epidemiology. *Am. J. Epidemiol.* **167,** 492–499 (2008).
198. Liu, J., Weinhandl, E. D., Gilbertson, D. T., Collins, A. J. & St Peter, W. L. Issues regarding 'immortal time' in the analysis of the treatment effects in observational studies. *Kidney Int.* **81,** 341–350 (2012).
199. Oksanen, A. Haemangiosarcoma in dogs. *J. Comp. Pathol.* **88,** 585–595 (1978).
200. Brown, N. O., Patnaik, A. K. & MacEwen, E. G. Canine hemangiosarcoma: retrospective analysis of 104 cases. *J. Am. Vet. Med. Assoc.* **186,** 56–58 (1985).
201. Wendelburg, K. M. *et al.* Survival time of dogs with splenic hemangiosarcoma treated by splenectomy with or without adjuvant chemotherapy: 208 cases (2001-2012). *J. Am. Vet. Med. Assoc.* **247,** 393–403 (2015).
202. Fosmire, S. P. *et al.* Canine malignant hemangiosarcoma as a model of primitive angiogenic endothelium. *Lab. Invest.* **84,** 562–572 (2004).
203. Maddox, J. C. & Evans, H. L. Angiosarcoma of skin and soft tissue: a study of forty-four cases. *Cancer* **48,** 1907–1921 (1981).
204. Daugaard, S., Hultberg, B. M., Hou-Jensen, K. & Mouridsen, H. T. Clinical features of malignant haemangiopericytomas and haemangioendotheliosarcomas. *Acta Oncol.* **27,** 209–213 (1988).
205. Karpeh, M. S., Jr, Caldwell, C., Gaynor, J. J., Hajdu, S. I. & Brennan, M. F. Vascular soft-tissue sarcomas. An analysis of tumor-related mortality. *Arch. Surg.* **126,** 1474–1481 (1991).
206. Benjamin, S. A., Hahn, F. F., Chieffelle, T. L., Boecker, B. B. & Hobbs, C. H. Occurrence of hemangiosarcomas in beagles with internally deposited radionuclides. *Cancer Res.* **35,** 1745–1755 (1975).
207. Nikula, K. J., Benjamin, S. A., Angleton, G. M., Saunders, W. J. & Lee, A. C. Ultraviolet radiation, solar dermatosis, and cutaneous neoplasia in beagle dogs. *Radiat. Res.* **129,** 11–18 (1992).
208. Ware, W. A. & Hopper, D. L. Cardiac tumors in dogs: 1982-1995. *J. Vet. Intern. Med.* **13,** 95–103 (1999).
209. Prymak, C., McKee, L. J., Goldschmidt, M. H. & Glickman, L. T. Epidemiologic, clinical, pathologic, and prognostic characteristics of splenic hemangiosarcoma and splenic hematoma in dogs: 217 cases (1985). *J. Am. Vet. Med. Assoc.* **193,** 706–712 (1988).
210. Falk, H., Thomas, L. B., Popper, H. & Ishak, K. G. Hepatic angiosarcoma associated with androgenic-anabolic steroids. *Lancet* **2,** 1120–1123 (1979).
211. Falk, H. *et al.* Epidemiology of hepatic angiosarcoma in the United States: 1964-1974. *Environ. Health Perspect.* **41,** 107–113 (1981).
212. Shi, E. C., Fischer, A., Crouch, R. & Ham, J. M. Possible association of angiosarcoma with oral contraceptive agents. *Med. J. Aust.* **1,** 473–474 (1981).

213. Lamerato-Kozicki, A. R., Helm, K. M., Jubala, C. M., Cutter, G. C. & Modiano, J. F. Canine hemangiosarcoma originates from hematopoietic precursors with potential for endothelial differentiation. *Exp. Hematol.* **34,** 870–878 (2006).
214. Yonemaru, K., Sakai, H., Murakami, M., Yanai, T. & Masegi, T. Expression of vascular endothelial growth factor, basic fibroblast growth factor, and their receptors (flt-1, flk-1, and flg-1) in canine vascular tumors. *Vet. Pathol.* **43,** 971–980 (2006).
215. Kato, Y. *et al.* Gene expressions of canine angiopoietin-1 and -2 in normal tissues and spontaneous tumours. *Res. Vet. Sci.* **81,** 280–286 (2006).
216. Murakami, M. *et al.* Expression of the anti-apoptotic factors Bcl-2 and survivin in canine vascular tumours. *J. Comp. Pathol.* **139,** 1–7 (2008).
217. Kim, J.-H. *et al.* Interleukin-8 promotes canine hemangiosarcoma growth by regulating the tumor microenvironment. *Exp. Cell Res.* **323,** 155–164 (2014).
218. Petterino, C., Rossetti, E. & Drigo, M. Immunodetection of the signal transducer and activator of transcription-3 in canine haemangioma and haemangiosarcoma. *Res. Vet. Sci.* **80,** 186–188 (2006).
219. Fukumoto, S. *et al.* Therapeutic potential of endothelin inhibitors in canine hemangiosarcoma. *Life Sci.* **159,** 55–60 (2016).
220. Thomas, R. *et al.* Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res.* **22,** 305–319 (2014).
221. Dickerson, E. B. *et al.* Mutations of phosphatase and tensin homolog deleted from chromosome 10 in canine hemangiosarcoma. *Vet. Pathol.* **42,** 618–632 (2005).
222. Heller, D. A. *et al.* Assessment of cyclooxygenase-2 expression in canine hemangiosarcoma, histiocytic sarcoma, and mast cell tumor. *Vet. Pathol.* **42,** 350–353 (2005).
223. Tamburini, B. A. *et al.* Gene expression profiling identifies inflammation and angiogenesis as distinguishing features of canine hemangiosarcoma. *BMC Cancer* **10,** 619 (2010).
224. Italiano, A. *et al.* Alterations of the p53 and PIK3CA/AKT/mTOR pathways in angiosarcomas: a pattern distinct from other sarcomas with complex genomics. *Cancer* **118,** 5878–5887 (2012).
225. Murai, A. *et al.* Constitutive phosphorylation of the mTORC2/Akt/4E-BP1 pathway in newly derived canine hemangiosarcoma cell lines. *BMC Vet. Res.* **8,** 128 (2012).
226. Behjati, S. *et al.* Recurrent PTPRB and PLCG1 mutations in angiosarcoma. *Nat. Genet.* **46,** 376–379 (2014).
227. Calvete, O. *et al.* A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li–Fraumeni-like families. *Nat. Commun.* **6,** 8383 (2015).

228. Murali, R. *et al.* Targeted massively parallel sequencing of angiosarcomas reveals frequent activation of the mitogen activated protein kinase pathway. *Oncotarget* **6,** 36041–36052 (2015).
229. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45,** D777–D783 (2017).
230. Wang, G. *et al.* Actionable mutations in canine hemangiosarcoma. *PLoS One* **12,** e0188667 (2017).
231. Iguchi, Y., Katsuno, M., Ikenaka, K., Ishigaki, S. & Sobue, G. Amyotrophic lateral sclerosis: an update on recent genetic insights. *J. Neurol.* **260,** 2917–2927 (2013).
232. van Blitterswijk, M. *et al.* Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **21,** 3776–3784 (2012).
233. Chiò, A. *et al.* Extensive genetics of ALS: a population-based study in Italy. *Neurology* **79,** 1983–1989 (2012).
234. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362,** 59–62 (1993).
235. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
236. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
237. Kabashi, E. *et al.* TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40,** 572–574 (2008).
238. Kwiatkowski, T. J., Jr *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323,** 1205–1208 (2009).
239. Vance, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323,** 1208–1211 (2009).
240. Maruyama, H. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465,** 223–226 (2010).
241. Deng, H.-X. *et al.* Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477,** 211–215 (2011).
242. Zeng, R. *et al.* Breed distribution of SOD1 alleles previously associated with canine degenerative myelopathy. *J. Vet. Intern. Med.* **28,** 515–521 (2014).
243. Crisp, M. J., Beckett, J., Coates, J. R. & Miller, T. M. Canine degenerative myelopathy: biochemical characterization of superoxide dismutase 1 in the first naturally occurring non-human amyotrophic lateral sclerosis model. *Exp. Neurol.* **248,** 1–9 (2013).

244. Carrillo, C. *et al.* Diacylglycerol-containing oleic acid induces increases in [Ca2+]i via TRPC3/6 channels in human T-cells. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1821,** 618–626 (2012).
245. Tseng, P.-H. *et al.* The canonical transient receptor potential 6 channel as a putative phosphatidylinositol 3,4,5-trisphosphate-sensitive calcium entry system. *Biochemistry* **43,** 11701–11708 (2004).
246. Marión, R. M., Schotta, G., Ortega, S. & Blasco, M. A. Suv4-20h abrogation enhances telomere elongation during reprogramming and confers a higher tumorigenic potential to iPS cells. *PLoS One* **6,** e25680 (2011).
247. Benetti, R. *et al.* Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J. Cell Biol.* **178,** 925–936 (2007).
248. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6,** 80–92 (2012).
249. Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22,** 1589–1598 (2012).
250. Fruman, D. A. *et al.* The PI3K Pathway in Human Disease. *Cell* **170,** 605–635 (2017).
251. Cully, M., You, H., Levine, A. J. & Mak, T. W. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nat. Rev. Cancer* **6,** 184–192 (2006).
252. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44,** 821–824 (2012).
253. Hoeppner, M. P. *et al.* An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* **9,** e91172 (2014).
254. Bloch, D. B. *et al.* Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. *Mol. Cell. Biol.* **20,** 6138–6146 (2000).
255. Lallemand-Breitenbach, V. de The H (2010) PML nuclear bodies. *Cold Spring Harb. Perspect. Biol.* **2,** a000661
256. Weber, E. W. *et al.* TRPC6 is the endothelial calcium channel that regulates leukocyte transendothelial migration during the inflammatory response. *J. Exp. Med.* **212,** 1883–1899 (2015).
257. Brazzatti, J. A. *et al.* Differential roles for the p101 and p84 regulatory subunits of PI3Kγ in tumor growth and metastasis. *Oncogene* **31,** 2350–2361 (2012).
258. Wang, X., Ding, J. & Meng, L.-H. PI3K isoform-selective inhibitors: next-generation targeted cancer therapies. *Acta Pharmacol. Sin.* **36,** 1170–1176 (2015).
259. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6,** 11 (2013).

260. Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y. & Dikic, I. Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* **416,** 183–187 (2002).
261. Yao, Y. *et al.* MiR-330-mediated regulation of SH3GL2 expression enhances malignant behaviors of glioblastoma stem cells by activating ERK and PI3K/AKT signaling pathways. *PLoS One* **9,** e95060 (2014).
262. Decker, B. *et al.* Homologous Mutation to Human BRAF V600E Is Common in Naturally Occurring Canine Bladder Cancer--Evidence for a Relevant Model System and Urine-Based Diagnostic Test. *Mol. Cancer Res.* **13,** 993–1002 (2015).
263. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417,** 949–954 (2002).
264. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
265. Komine, O. & Yamanaka, K. Neuroinflammation in motor neuron disease. *Nagoya J. Med. Sci.* **77,** 537–549 (2015).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1454

Editor: The Dean of the Faculty of Medicine

Distribution: publications.uu.se
urn:nbn:se:uu:diva-347165

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018