



UPPSALA  
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 1624*

# Characterizing the spectrum of somatic alterations in canine and human cancers

SHARADHA SAKTHIKUMAR



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2020

ISSN 1651-6206  
ISBN 978-91-513-0839-5  
urn:nbn:se:uu:diva-399978

Dissertation presented at Uppsala University to be publicly examined in B41, Biomedicinskt centrum (BMC), Husargatan 3, Uppsala, Thursday, 20 February 2020 at 09:14 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Richard Rosenquist Brandell.

## Abstract

Sakthikumar, S. 2020. Characterizing the spectrum of somatic alterations in canine and human cancers. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1624. 64 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0839-5.

Cancers arise as a result of deleterious somatic alterations accumulating in the genome during the process of cell division. These alterations arise either via exposure to mutagens or due to errors occurring during DNA replication. In this thesis, a systematic exploration, from discovery to analyses of somatic alterations in three diverse cancers that affect dogs and humans, was undertaken.

In Studies I and II, whole-exome sequencing of dogs affected by the cancers of osteosarcoma and hemangiosarcoma were done to delineate coding mutations that can contribute to their carcinogenesis. Besides, as these cancers mirror the corresponding human disease in clinical manifestation and histological features, a secondary objective was to confirm the molecular drivers found in the canines were also influencing factors in the human cancer(s).

In the osteosarcoma investigations with three breeds, we found that tumors show a high frequency of somatic copy-number alterations, affecting key cancer genes. *TP53* was the most frequently altered gene, akin to human osteosarcoma. The second most mutated gene, histone methyltransferase *SETD2*, has known epigenetic roles in multiple cancers but not in osteosarcoma. Our study highlights the strong genetic similarities between human and dog osteosarcoma, suggesting that canine disease may serve as an excellent model for developing treatment strategies in both species.

In the hemangiosarcoma study in golden retrievers, putative driver alterations were identified in the tumor suppressor *TP53* and in genes involved in the cell cycle regulating PI3K pathway, including *PIK3CA* and *PIK3RI*. Furthermore, we find several somatic alterations between the dog hemangiosarcoma and human angiosarcoma overlap, indicating we can use the canine model to apprise the infrequently occurring human disease.

In Study III, we implemented whole-genome sequencing methodologies to define both coding and non-coding alterations in the glioblastoma cancer genome. We find the coding somatic alterations recapitulate what has been previously seen for the cancer, including driver alterations in the genes of *EGFR*, *PTEN*, and *TP53*. Significantly though, using the concept of evolutionary constraint, we find an enrichment of non-coding mutations in regulatory regions, around GBM-implicated genes. The mutated regions include splice sites, promoters, and transcription factor binding sites, suggesting the importance of regulatory mutations for the pathogenesis of glioblastoma.

Overall, the insights garnered from the above exome- and genome-wide surveys provide novel insights into unraveling some of the complexities associated with somatic genomic alterations in cancer genomes. It also convincingly underscores the benefits of using sequencing technologies to comprehend complex biological diseases.

**Keywords:** dog, osteosarcoma, hemangiosarcoma, glioblastoma, non-coding, whole-exome, whole-genome, sequencing, bioinformatics, comparative genetics.

*Sharadha Sakthikumar, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden. Science for Life Laboratory, SciLifeLab, Box 256, Uppsala University, SE-75105 Uppsala, Sweden.*

© Sharadha Sakthikumar 2020

ISSN 1651-6206

ISBN 978-91-513-0839-5

urn:nbn:se:uu:diva-399978 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-399978>)

*For Sakthi Appa, TNA Thatha, and  
Amma Vatsala*



“Strange things may be generally accounted for if their cause be fairly searched out.”

—*Jane Austen, Northanger Abbey*



# List of Papers

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

- I **Sakthikumar S**, Elvers I, Kim J, Arendt ML, Thomas R, Turner-Maier J, Swofford R, Johnson J, Schumacher SE, Alföldi J, Axelsson E, Couto CG, Kisseberth WC, Pettersson ME, Getz G, Meadows JRS, Modiano JF, Breen M, Kierczak M, Forsberg-Nilsson K, Marinescu VD, Lindblad-Toh K. ***SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma.*** *Cancer Res.*, 2018 78 (13):3421-3431.
- II Megquier K, Turner-Maier J, Swofford R, Kim J, Sarver AL, Wang C, **Sakthikumar S**, Johnson J, Koltoonian M, Lewellen M, Scott MC, Schulte AJ, Borst L, Tonomura N, Alföldi J, Painter C, Thomas R, Karlsson EK, Breen M, Modiano JF, Elvers I, Lindblad-Toh K. ***Comparative Genomics Reveals Shared Mutational Landscape in Canine Hemangiosarcoma and Human Angiosarcoma.*** *Mol. Cancer Res.* 2019 17(12):2410-2421.
- III **Sakthikumar S\***, Roy A\*, Haseeb L\*, Pettersson ME, Marinescu VD, Lindblad-Toh K, Forsberg-Nilsson K. ***Whole Genome Sequencing of Glioblastoma Reveals Enrichment of Non-Coding Constraint Mutations in Known and Novel Genes.*** Manuscript in revision.

*\*Authors contributed equally.*

Reprints were made with permission from the respective publishers.





# Contents

1	Introduction .....	13
1.1	What is cancer, and how does it develop? .....	13
1.2	Somatic mutation theory of cancer, tumor suppressors and oncogenes .....	14
1.3	Strategies for studying cancer .....	16
1.4	Massively parallel sequencing and its application in cancer .....	17
1.5	The dog as a model organism for the study of complex diseases ..	18
1.6	Comparative genomics, conservation, and their application to understanding diseases .....	21
2	Overview of the cancers investigated .....	23
2.1	Types of cancer .....	23
2.2	Osteosarcoma .....	23
2.3	Hemangiosarcoma and Angiosarcoma .....	25
2.4	Glioblastoma .....	26
2.5	In summary .....	28
3	The current studies .....	29
3.1	Aims of the thesis .....	29
4	Remarks on methods implemented in the current studies .....	30
4.1	Overview of cancer genome analysis .....	30
4.2	Whole-exome and whole-genome sequencing of cancer .....	31
4.3	Bioinformatics analyses of cancer genomes in the current studies	33
4.4	Statistical analysis employed .....	35
5	Summary of current studies .....	36
5.1	SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma (Paper I) .....	36
5.1.1	Background .....	36
5.1.2	Results .....	36
5.1.3	Discussion .....	37
5.2	Comparative Genomics Reveals Shared Mutational Landscape in Canine Hemangiosarcoma and Human Angiosarcoma (Paper II) .....	38
5.2.1	Background .....	38
5.2.2	Results .....	38
5.2.3	Discussion .....	39

5.3	Whole genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes (Paper III)	40
5.3.1	<i>Background</i>	40
5.3.2	<i>Results</i>	40
5.3.3	<i>Discussion</i>	43
6	Concluding remarks and future directions	45
6.1	Overview	45
6.2	Limitations of the current investigations	46
6.3	Ongoing work and future directions	48
7	Acknowledgements	50
8	References	52

# Abbreviations

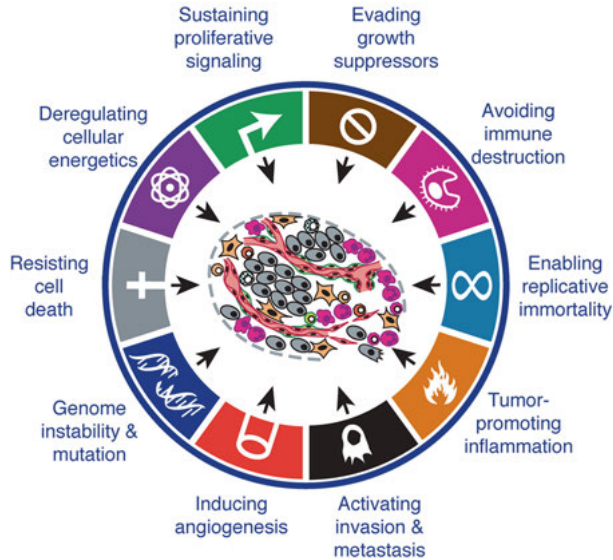
AS	angiosarcoma
bp	base pair
CNS	central nervous system
CNV	copy number variations
DoC	depth of coverage
FDR	false discovery rate
GBM	glioblastoma
GH	greyhounds
GR	golden retrievers
GWAS	genome-wide association studies
HSA	hemangiosarcoma
kbp	kilobase pair
LSA	lymphoma
Mbp	megabase pairs
MPS	massively parallel sequencing
NCCMs	non-coding constraint mutations
NGS	next-generation sequencing
NMF	nonnegative matrix factorization
OSA	osteosarcoma
RW	Rottweilers
SCNA	somatic copy number alterations
SIM	somatic indel mutations
SMG	significantly mutated genes
SNP	single nucleotide polymorphisms
SPM	somatic point mutations
SV	structural variants
TF	transcription factors
TFBS	transcription factor binding sites
TSG	tumor suppressor gene
VAF	variant allele fraction
WES	whole-exome sequencing
WGS	whole-genome sequencing



# 1 Introduction

## 1.1 What is cancer, and how does it develop?

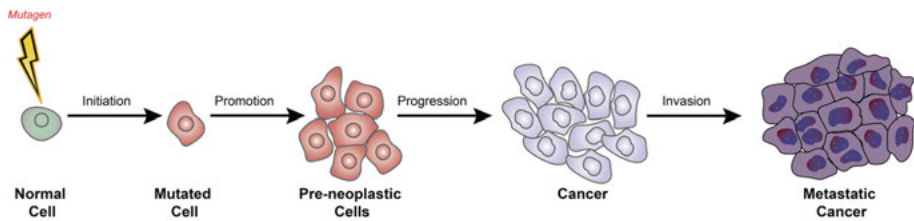
Cancer, ‘The Emperor of all Maladies’ [1], is *not* one disease, but in actual fact represents a group of neoplastic disorders that are characterized by genetically unstable cells that are capable of dividing uncontrollably and grow in an atypical manner. In the year 2000, Hanahan and Weinberg, in their seminal paper “Hallmarks of Cancer” [2], describe how normal cells, via a multi-step process, acquire biological traits or functional capabilities that make them turn cancerous. The authors stipulate that the important characteristics for a normal cell to become cancerous include development of: i) the ability to grow in the absence of growth stimulatory signals; ii) insensitivity to inhibitory signals that stop unregulated growth; iii) the capability to elude apoptosis (programmed cell death); iv) indefinite proliferative capacity; v) the ability to stimulate angiogenesis, which can result in the formation of new blood vessels that can deliver oxygen and essential nutrients to the cancer cells; vi) the competence to invade neighboring tissues, enter the blood or lymphatic systems, and spread to other parts of the body. The authors [3] have since amended the list of hallmarks to include two more capabilities that cancer cells must possess: vii) the ability to reprogram the energy metabolism; viii) and lastly, the capacity to evade immune suppression/destruction. More importantly, the authors aver that there are two enabling aberrations that facilitate the normal cell’s acquisition of the above functional capabilities: genome instability and tumor-promoting inflammation. The eight traits and the two enabling characteristics together are described as “The Hallmarks of Cancer: Next Generation” (**Figure 1**).



**Figure 1.** *The Hallmarks of Cancer. The schema elucidates the necessary criteria/conditions for the manifestation of cancer. Reproduced with permission from Hanahan and Weinberg [3].*

## 1.2 Somatic mutation theory of cancer, tumor suppressors and oncogenes

Historically, multiple theories have been advanced to explain the incidence of cancer [4]. One of the earliest tenable ideas was put forth by Theodor Boveri in 1914, wherein he believed that groups of chromosomal aberrations in vulnerable cells lead to carcinogenesis [5]. The notion was reinforced soon after by Karl-Heinrich Bauer’s proposal that mutations in cells are responsible for the formation of cancer [6]. Boveri’s model, now referred to as the “somatic mutation theory” (SMT; **Figure 2**), acquired traction when, in 1953, the Norwegian statistician Carl Nordling published “A new theory on cancer-inducing mechanism” [7]. Here, he conjectured that over time, somatic cells can acquire gene mutations, and when many genes become mutated, they can cause cells to form cancer.



**Figure 2.** According to the SMT, cancer is the result of successive mutations. A normal cell acquires a somatic alteration when exposed to an endogenous or exogenous mutagen, which can then initiate and go through the phases of carcinogenesis as shown. Adapted with permission from Siddiqui et al. [8].

In 1969, D.J. Ashley [9] estimated that the process from initiation to formation of cancer is the result of a discrete/finite number of mutations, a theory corroborated in 1971 by Alfred Knudson, with his two-hit hypothesis [10]. When Knudson compared the incidence of inherited versus sporadic (non-heritable) forms of retinoblastoma, he found that the former required only one mutation, whereas the latter needed two. He found that patients who were afflicted with the heritable variety had a susceptibility or germline mutation in the negative regulator of the cell cycle gene Retinoblastoma 1, *RBI*. Only one mutation in the somatic cell was needed for carcinogenesis to occur. Conversely, for the sporadic type, two mutations were required in the gene, leading to its ‘loss-of-function’ and, consequently, the manifestation of the disease.

*RBI* was the first tumor suppressor gene (TSG) to be identified [11]. This class of genes helps regulate cell growth, in particular inhibit uncontrolled growth. When inactivated by mutation, it can lead to the formation of cancer [12]. TSGs are typically recessive at the cellular level, and usually in cancers both alleles of the gene get inactivated [13]. Examples of TSGs include *TP53*, which is seen to be mutated in >50% of all human cancers [14]; *BRCA1/2* are two genes where mutations in both the germline as well as somatic cells are known to have clinical consequences for breast cancer patients [15, 16]. *PTEN* is another TSG that is seen across several cancers; its loss of function occurs through either point mutations, copy number alterations or transcriptional silencing of the gene [17].

Another category of genes consists of the so-called oncogenes; these originate from proto-oncogenes, genes that help in the regulation of cell growth and enhance cell proliferation [13]. However, when proto-oncogenes get active through ‘gain-of-function’ mutations, it can fuel uncontrolled cell growth and division, thereby triggering carcinogenesis [13]. Examples include the *ras* family of proto-oncogenes (*H-ras*, *K-ras*, and *N-ras*), and mutated forms of these are found across several cancers, including carcinomas and myelodysplastic syndromes [18-20]. The *MYC* family of oncogenes are also seen to be dysregulated in several cancers, and this dysregulation is often associated with poor clinical outcomes [21]. Oncogenes are dominant, and unlike TSGs, they only need one mutation in one of the two alleles for the initiation of cancer. In

*vivo* studies in animals [22-24], where mutated oncogenes were inserted into healthy cells leading to their carcinogenic transformation, intriguingly serve to underscore SMT, *i.e.*, cancer starts with a mutation, followed by clonal expansion, the appearance of the disease, and eventual metastasis.

### 1.3 Strategies for studying cancer

Cancer affects one in five men and one in six women [25] and is the second leading cause of death globally [26]. In 2018, there were an estimated ~18 million cases worldwide and ~9 million deaths as a result of the disease [25]. Not only does cancer have tragic consequences for the affected individuals and their families, but it also is a heavy burden on the taxpayer. For example, the projected cancer-associated expenditure in the United States for the year 2020 is USD 173 billion [27]. Cancer is predominantly a disease of the aging [28], and with life expectancy increasing worldwide [29], the incidence of the disease and its associated mortality are expected to be on the rise, along with the economic burden, which is also projected to increase considerably following diagnosis [30]. Therefore, concerted research efforts in allied fields are necessary, not only to understand the biological mechanisms behind the disease, but also to create new therapeutic strategies and eventually find cures for this dreadful disease.

There are many systematic approaches for investigating cancer; traditional methods include histopathological and cytogenetic studies. The former method is the most prevalent, and involves studying alterations in cancer tissues, while the latter entails probing the disease in terms of chromosomal counts and possible modifications to their structure. These strategies fall under the framework of what can be termed as ‘organ-focused’ approaches. However, cancer is inherently a disease of genes/genome malfunctioning and possibly epigenetic dysregulation, and the above methodologies shed very little light on the genetic origins of the disease. It is imperative that there be a shift or a complementary approach towards investigating the disease at the genomic level to better apprehend the associated pathogenesis.

The advent of human genome sequencing [31, 32], targeted gene sequencing panels [33-35] and genome-wide association studies (GWAS) for the detection of inherited mutations [36], and the application of the so-called next-generation sequencing (NGS) technologies encompassing the gamut of whole-genome sequencing (WGS) and whole-exome sequencing (WES) to study tumor mutations, and/or—transcriptome sequencing (RNA-seq) to study changes in gene expression [37-40], have endowed researchers with disease insights never before gleaned. These approaches give us effective ways to look at associations between genetic variants and disease traits, cancer genome alterations—be they base-pair level changes, insertions/deletions (indels), copy



number alterations, or chromosomal rearrangements—and garner an improved understanding of the underlying disease biology.

For example, investigations for germline predisposing factors in hereditary cancers can be done using the GWAS methodology. This entails scanning for markers, often in the form of single nucleotide polymorphisms or SNPs across the genomes of many individuals in a population to find inherited genetic variations that may be associated with a specific disease [41]. From over 350 cancer GWAS studies, more than 2,500 unique SNPs were shown to be associated with solid and hematological tumors [42].

Although germline mutations predispose an individual to cancer, it is the accrual of additional somatic alterations in the genome that is likely to be the causal factor for malignancy. To discern these alterations (which could be either point mutations or larger structural changes), targeted or WGS methods are employed [43]. Systematic high-throughput sequencing analysis has led to the discovery and validation of mutations in genes such as *AKT* [44], *EGFR* [45], and *BRAF* [46], all of which are known to have roles in the somatic progression of cancers. The next section elaborates further on sequencing strategies and their applications.

## 1.4 Massively parallel sequencing and its application in cancer

The introduction of Sanger sequencing, based on capillary electrophoresis in the mid-1970s, brought about a revolution in the field of genomics that culminated in the sequencing of the human genome in the 1990s. Sanger sequencing (also referred to as first-generation sequencing) as yet is the gold standard for obtaining contiguous DNA fragments. The fragments with per-base accuracies as high as 99.999% are typically 600-700 base pair (bp) long, though they can be sequenced up to ~1 kilobase pair (kbp) length [47]. Notwithstanding these advantages, the Sanger methodology is prohibitively expensive, of relatively low throughput, and can only be done in large sequencing centers and by consortiums. In 2005, the appearance of the massively parallel sequencing (MPS) methods ushered in the era of next-generation sequencing [47]. Fundamentally, NGS technologies are akin to Sanger sequencing, wherein DNA fragments are sequenced, albeit to smaller lengths; in contrast, though, with NGS's sequencing chemistry, it is possible to sequence millions of DNA fragments in parallel and achieve a high level of redundancy or multiple representations of a nucleotide or base, which allows for its accurate characterization [48]. Commercial vendors that offer NGS sequencing include Illumina, for short-read sequencing, PacBio, the most widely used for long-reads, and Oxford Nanopore, which offers sequencing of both the forward and reverse strands of DNA [49, 50]. NGS, in conjunction with the mandatory bioinformatics-based methods, has enabled researchers to perform genomic analyses

in a systematic and high-throughput manner, which has resulted in accelerating the pace of biological and biomedical investigations. The technology has wide-ranging applications, including genome re-sequencing, variation detection, gene discovery, gene expression profiling with RNA sequencing, DNA methylation, and epigenetics [51].

As previously noted, from a molecular standpoint, cancer is a disease that involves malignant or harmful alterations to the DNA, and the ability to detect and decode the changes wreaked here is fundamental to understanding the disease's genetic origins and related mechanisms. The cancer genome can now be explored in high resolution and with sensitivity to discern alterations that initiate and drive tumor development through the NGS' WES, WGS or transcriptome sequencing approaches [35, 52]. Although it is possible to do a *de novo* detection of tumor changes, the majority of the cancer studies thus far employ matched paired investigations, in which both the normal and tumor genomes of the same individual (normal DNA coming most often from blood, alternatively from tissues near the tumor) are sequenced to discriminate somatic mutations from germline mutations [53]. In cancer genome analysis, it is critical to have matched pairs sequenced to enough depths of coverage (DoC; the number of times a base is covered during sequencing). Sequencing depth influences the sensitivity and specificity of both germline and somatic variant detection. The higher the DoC, the greater the probability of calling a mutated base and distinguishing it from sequencing errors [54].

Furthermore, if the somatic mutations exist only in a subset of the tumor cells (for example, only in subclones), it is critical to sequence to higher depths of coverage to be able to delineate these variants correctly. Studies have shown that for accurate and confident somatic variant detection, the normals and tumors be sequenced to minimum depths of coverage of 30X and 60X, respectively [55, 56].

So far, several studies have exploited NGS to characterize genome-wide molecular alterations in diverse cancers, including breast [57], lung [58], cervical [59], colorectal [60] and multiple myeloma [61]. However, this is only the "tip of the iceberg." With the dawn of third-generation sequencing technologies, our knowledge and understanding of the cancer genome will be further enhanced, wherein we will be able to differentiate intricate genomic aberrations, detect multiple transcript isoforms, epigenomic modifications, and their phase statuses [62]. These efforts, it is hoped, will lead to improved diagnosis, better cure, and ultimately actual prevention of the disease.

## 1.5 The dog as a model organism for the study of complex diseases

After the sequencing of the human genome, high-quality assemblies for the genomes of numerous model organisms, including the mouse and dog, have been generated [63, 64]. Comparative genomics approaches have shown that

humans share most of their genes and molecular pathways with some of these above model organisms [64, 65]. For example, dogs have approximately 20,257 genes, which is of the same order of magnitude as seen in humans (20,805; [www.ensembl.org](http://www.ensembl.org)), and most of these genes are known to be orthologous. Therefore it is probable that they will carry out the same essential biological functions, including being the possible genetic basis for similar disease phenotypes across the two species. Hence genetic insights gathered from dogs can be used for further understanding of human biology, including the origin of diseases [65].

*Canis familiaris*, the domestic dog, is an invaluable model organism in biomedical research for several reasons. Dogs are easily accessible and have an important and prized status in homes across diverse societies. In the United States, an excess of USD 40 billion is spent annually on health care for dogs [66]. As a consequence, the shared environment of pet owners and dogs can be taken advantage of, to study the epidemiology and genetics of diseases that are common to both dogs and humans. Of all the mammals, excepting humans, dogs have the most phenotypic diversity and the broadest spectrum of naturally occurring diseases [67]. On a phylogenetic tree, dogs are more distant from humans than the commonly used model organism, the mouse; however, dogs possess a higher degree of sequence and protein similarity with the human genome [64]. Hence many biological processes in dogs, which occur as a result of changes either at the regulatory or at the protein sequence level, are presumably conserved in humans as well. Nearly ~400 ailments related to those of humans are documented in dogs, including complex illnesses such as cancers, heart disease, and neurological disorders [68, 69]. Given the anatomical and physiological similarities between dogs and humans, disease manifestations and clinical responses in dogs often closely mirror human diseases, indicating that they might share similar genetic pathways [70]. In more than 40 conditions (monogenic as well as multi-factorial), dogs are known to have mutations in the genes homologous to human genes associated with the disease phenotype [71].

Dogs are prone to several cancers that strike humans also, including lymphoma, osteosarcoma, glioblastoma, bladder, and breast cancer [72-76]. These dog tumors share characteristics with human cancers, such as in their histology, etiology, the spontaneity of occurrence, and similar progression of the disease [77]. Furthermore, many types of cancers in both humans and dogs have similar genetic alterations, be it germline or somatic mutations or copy number alterations, among others. For instance, a study comparing the somatic copy number alterations (SCNA) and progression of colorectal cancer (CRC) in humans and dogs [78] showed, firstly, that the tumors in both species have similar SCNAs, and, secondly, that the same genetic pathways were affected by the genes in the aneuploidy regions. Lastly, when clustering analysis was done using overlapping SCNAs, it was observed that the samples clustered, not as expected at the species level, but according to the site of origin or the

stage of the disease. These findings indicate that dog-human recurrent SCNAs and the altered genes they contain are probable drivers that cause CRC in both species. (Driver mutations in cancer genomes contribute to malignant initiation or development).

In an investigation of the cancer lymphoma (LSA) in dogs, using WES of matched tumor and normal samples, Elvers *et al.* [79] discovered somatic mutations in multiple genes that are also found in human lymphomas. The study included three dog breeds—cocker spaniels (CS), golden retrievers (GR), and boxers (BX)—each with differing risk for lymphomas arising from either B- or T-cells. For the T- type LSA, there was no overlap among the significantly mutated genes (SMG) between BX and GR. In contrast, for the B-cell LSA, multiple SMGs overlapped between CS and GR; these include *TRAF3*, *MAP3K14*, *FBXW7*, and *POT1*. The *FBXW7* mutations recur at a specific codon, and the matching codon in humans is known to be frequently mutated in LSA. Thus the identification of genes that are common with the human disease and of novel risk-associated genes in dogs may increase our understanding of the LSA disease mechanism, and in turn, may lead to new therapeutic targets for both species.

Some heritable cancers in dogs are known to be similar to human familial cancer syndromes. Therefore, identifying and cataloging germline mutations in cancer-predisposing genes in the dog can potentially be used to identify regions in the human genome that harbor disease susceptibility loci [80]. A GWAS performed in GR dogs, a breed that has 6% and 20% predisposition to lymphoma and hemangiosarcoma respectively, uncovered a locus that confers risk for both malignancies [81]. (These two dog tumors are known to be equivalent to human B-cell non-Hodgkin lymphoma and angiosarcoma, correspondingly). The authors also identify three shared and one lymphoma-specific risk haplotypes within the two loci. Although no coding changes were linked to the risk haplotype, differential gene expression analysis of the haplotype identified over 100 genes that are associated with immune cell activation. Therefore it was suggested that the disposing germline variants in the risk loci are likely regulatory in nature and affect pathways that mediate T-cell regulation and proliferation. The authors thus conclude that the interface between the immune system and malignant tumor cells indicates a common role in the oncogenesis of these comparatively diverse cancers.

In addition to having similar diseases as humans, dogs have a population structure that has primarily contributed to the use of their genomes for studying the genetic causes of diseases. The dog population has undergone two significant bottlenecks, the first resulting from domestication more than 15,000 years ago [82], and the second due to either selection or genetic drift during intentional breed creation over the last 200 years [69]. The latter rigorous artificial selection, through which desirable phenotypes were selected from a few founders, has led to the creation of more than 400 distinct dog breeds with great diversity in morphological features and/or behavioral traits [83]. The

haplotype structure of the dog genome has been influenced by these two population bottlenecks. (Haplotype blocks/structures are long stretches of DNA along a chromosome and are often lacking in recombination events). Dogs from the initial domestication of wolves carried forward a limited gene pool in short haplotype blocks – the domestic dog has a haplotype block size across breeds of about 10-kbp and shorter than what is found in humans [84]. In contrast, selection for specific traits during the recent phase of breed creation has resulted in long haplotype blocks measuring up to 1-megabase pairs (Mbp) within breeds [64].

The recent inbreeding of dogs has also led to an enrichment of specific genetic variants, some of which may confer disease susceptibility in certain breeds. Several autoimmune diseases, behavioral syndromes, and cancers are known to be more prevalent in certain breeds over others [85]. Nearly all of these diseases have analogs in the human disease spectrum, and the segregation of diseases based on breeds suggests that the genetic background plays a factor. This fact can, therefore, be exploited to study diseases by focusing on one or more breeds.

## 1.6 Comparative genomics, conservation, and their application to understanding diseases

In the previous section, the extensive orthologous relationship between dogs and humans relative to the protein-coding part of the genome was discussed, and how it is probable that the corresponding encoded genes have the same biological functions cross-species. The field of comparative genomics, where comparison of genomes, akin to the above human-dog or other distantly related species, is made, has increased our knowledge of how genomes evolve. In addition, it has also aided in the identification of functional elements, both in the coding as well as non-coding (regulatory) part of the genomes [86], and has led to a better grasp of molecular mechanisms that are conserved across several taxa [87].

The foundation of the field of comparative genomics is based on the premise that essential biological sequences or functionally important residues among species are encoded in evolutionarily conserved DNA as a result of functional constraints [86]. Therefore it follows that specific genomic alterations in the functional constrained elements (regions of highest sequence conservation) are likely to disrupt function and be damaging to an organism [87].

The question that begs to be answered now is: how do we connect or associate genomic changes with a damaging phenotype that is observed? This question is especially important to resolve in diseases like cancer. Coding alterations that are responsible for carcinogenesis can be detected with relative ease. For a given cancer, it entails looking for signals of positive selection in

genes across several tumors, in which the genes tend to have higher mutation frequencies than what is observed in the background. The non-coding alteration that is detrimental or damaging to the cancer cannot be marked as straightforwardly, however. For one, there are substantially more mutations in the non-coding disease genome than what is found in the coding regions. (Studies have shown that >90% of disease-associated variants are likely to be in the non-coding regions of the genome [88]). Secondly, to prioritize the changes, we need to know if the mutations actually overlap functional sequences of the non-coding genome and are therefore likely to have a deleterious effect.

Delineating the functional sequences of the non-coding genome is challenging, but projects like the Encyclopedia of DNA Elements (ENCODE; [89]) have used the assays of DNA methylation, ChIP-seq, DNase I Hypersensitivity and RNA-seq to identify functional regions of the human genome. ENCODE information can be used to annotate regions of the genome, including regulatory elements that control cellular functions and also elements that function at the RNA level. A comparative analysis with 29 mammalian genomes by Lindblad-Toh and colleagues [90] revealed that >4% of the human genome contains constrained elements. And ~60% of the constrained bases likely overlap regulatory elements that include promoter, insulator, and enhancer regions. The authors contend that with the information available about constrained elements, it will be possible to annotate variants in a disease, and center on those that likely disrupt regulatory functions. ‘Genomic Evolutionary Rate Profiling’ (GERP; [91]), a tool devised for estimating evolutionary constraint based on sequence conservation across species, also discerns constrained elements, and GERP threshold scores can potentially be used to define regions that may contain putative functional elements. In the study of glioblastoma in this thesis, we have applied the concepts of sequence conservation and GERP scores to understand better the somatic alterations that occur in the non-coding genome, and this is further elaborated in Chapter 5.

## 2 Overview of the cancers investigated

### 2.1 Types of cancer

There are over 200 types of cancers, with different classification schemes for the disease based on the histological features, tissue of origin, differentiation states, and biological behavior of the associated tumor cell [13]. For instance, based on the cell type the tumor cells are originating from, cancers can be categorized as either as carcinomas (arising from epithelial cells), sarcomas (occurring in mesenchymal cells, e.g., bone or muscle cells), adenocarcinomas (developing from glandular tissue, e.g., breast), leukemias (arising from bone marrow stem cells), lymphomas (initiated in lymphocytes) or melanomas (evolving in melanocytes). There are also cancers that occur in the central nervous system (CNS), which affect the brain and spinal cord, such as glioblastoma, medulloblastoma, meningioma, and intramedullary tumors, among others.

In this thesis, the focus is on two sarcomas and one CNS tumor, the profiles of which are expanded on in the next three sections of the current chapter.

### 2.2 Osteosarcoma

Osteosarcoma (OSA) is the most prevalent type of primary cancer of the bone in humans [92]. It is an aggressive tumor that is diagnosed in children and adolescent patients, with a second peak in frequency observed in adults over the age of 50 [93]. For the childhood type of the disease, the incidence is 5.0 per million persons per year, regardless of ethnicity, whereas for the adult form, it is 4.6, 6.5, and 6.8 per year per million persons per year for Caucasians, Hispanics, and Blacks respectively [93]. OSA commonly occurs in the long bones of the extremities near the metaphyseal growth plates, and the most common affected sites are the femur, the tibia, and the humerus [93]. The five-year survival rate for osteosarcoma is 46% [94].

The majority of the OSA tumors are high grade and are genomically unstable tumors with complex disorganized karyotypes [95]. At the chromosome level, there is a high level of instability, with parts or entire chromosomes being duplicated or deleted, leading to high levels of somatic structural variations and copy number alterations [96, 97]. The molecular characterization of

the disease can, therefore, be impeded by the extensive genomic variability as well as by the presence of intra- and inter-tumoral heterogeneity in the tumors.

OSA in humans is mostly sporadic; however, if the gene *RBI* is inactivated through a germline mutation, the risk of developing osteosarcoma increases significantly [98]. Also, it has been shown that individuals who are susceptible to the autosomal recessive diseases of Werner, Bloom, and Rothmund-Thomson syndromes [99] have an increased predisposition to osteosarcoma [99].

In dogs, osteosarcoma accounts for approximately 85% of bone tumors [100], and the average incidence has been estimated to be ~14 per 100,000 dogs per year [101]. It mostly affects middle-aged dogs, around 7.5 years [102] of age; there is a second, smaller peak in which 6-8% of cases are <3 years of age [103]. The size of the dog is a stronger predictor than its age, however. Large (25-40 kilograms) and giant breeds (more than 40 kilograms) account for 55% and 35%, respectively, of all cases of OSA [102]. As in the case of humans, the primary tumor of the bone comprises malignant stroma, with osteoid formation [104]. The sites of the tumor also reflect the human counterpart appendages, including the tendency of having more tumors in the appendicular versus the axial regions [105]. Approximately 90% of the tumors in humans and ~80% in dogs are appendicular tumors. Unlike in humans, dog breeds cover a span of values for relative risk and mortality from the disease, indicative of an apparent genetic predisposition for the disease among certain breeds over others [106]. For example, among the breeds of GR, Rottweiler (RW) and greyhound (GH), the incidence rates per 10,000 dog-years at risk (DYAR) are 6.0, 36, and 30 respectively [107].

Karlsson *et al.* [108], through a GWAS with the dog breeds of GH, RW, and Irish wolfhounds, discovered 33 inherited risk loci for OSA predisposition. These loci explain 50-80% of disease risk for the breeds, and the top locus was found to be a non-coding region on chromosome 11 near the tumor suppressor genes of *CDKN2A/2B*. The risk haplotype 11q16 is syntenic to a regulatory region on human chromosome 9p21 and is known to be deleted in 5% to 21% of human OSA cases [109]. This region of chromosome 9 is also one of the most complex regulatory areas in the human genome. It was therefore postulated that variants in the risk haplotype might disrupt enhancer elements upstream of the *CDKN2A/B* locus, thus altering the expression of genes in the region. Functional enrichment analysis of the genes in the 33 loci implicate pathways that are linked with bone growth and formation.

Despite discerning genetic alterations in both human and dog OSA, few recurrent clinically actionable mutations have been found. For around 35 years up to the present day, it is standard practice for OSA patients from both species to be treated with surgery, followed by rigorous adjuvant chemotherapy [110, 111]. For human OSA, immunotherapy with denosumab, a monoclonal antibody, is currently being tested in clinical trials to verify if it can inhibit the growth and spreading of the tumor cells (<https://clinicaltrials.gov/ct2/show/study/NCT02470091>). There is also an indication in pre-



clinical studies that immune checkpoint inhibitors can be used in the management of the disease [112].

## 2.3 Hemangiosarcoma and Angiosarcoma

Hemangiosarcoma (HSA) is an extremely invasive and malignant canine tumor that occurs in the vascular endothelium or endothelial precursor cells [100, 113]. There is a milder form induced by ultraviolet radiation, which occurs as a cutaneous mass that can be treated [114]. In contrast, though, the splenic type of HSA is often described as being a ‘silent’ killer since the clinical manifestation is almost indiscernible until after an acute collapse following the rupturing of the primary tumor [100].

HSA is a relatively common cancer in dogs, with several breeds at risk, including boxers, bulldogs, golden retrievers, Labrador retrievers, and Great Danes, among others [114]. The median survival period for the affected dogs is dismal, and most die within six months of diagnosis [115]. There are no standard therapies offered so far; doxorubicin, an anthracycline, is a moderately effective chemotherapeutic drug, though survival following its administration is only nominal [116].

The etiology of the disease is mostly unknown; however, given that there are a few select breeds that are more prone to the disease than others, it indicates that heritable predisposing factors are enabling the formation of the tumor [117]. The molecular basis for HSA also has not been clearly delineated. However, in a GWAS study [81] with golden retrievers from the USA, which carry a 20% lifetime risk for HSA, the authors identified two loci on chromosome 5 that contribute to an ~20% risk for contracting the disease. Whole-genome resequencing and subsequent genotyping identified several risk haplotypes in the two loci. The authors also find that none of the predisposing germline mutations were coding; instead, they appear to be regulatory in nature. In another whole-exome sequencing study with 20 affected dogs, somatic mutations in known cancer genes, including *TP53*, *PTEN*, and *PIK3CA*, were seen in >50% of the dogs [118].

Angiosarcoma (AS), the human histological equivalent of canine hemangiosarcoma, is a rare and fatal cancer [119]. The disease appears to stem from cells that are endothelial in nature, analogous to what is observed in dogs [119]. Though the disease can occur anywhere in the body, the most common manifestation is in the form of cutaneous diseases in the head and neck regions in Caucasian men over the age of 70 [119, 120]. The five-year survival rate for non-metastatic disease is ~45% [121]. One of the earliest known causes for the disease was attributable to therapeutic radiation with the commercial drug ‘Thorotrast’, which contained radioactive thorium [122]. Another plausible reason for the disease is exposure to vinyl chloride [123, 124], though not all in contact with the chemical agent are affected.

Because of the lack of human samples for studying the genetic basis of the disease, the molecular pathology remains largely unknown. However, a few cytogenetic studies of radiotherapy-induced AS indicate that it is an aneuploid cancer with amplifications seen in the *MYC* locus [125, 126]. Gains have also been observed for the genes of *FLT4* and *VEGFR3* [126, 127]. Applying a combination of WGS and WES approaches with 39 tumors, Behjati *et al.* identified frequent mutations in the angiogenesis signaling pathway genes of *PTPRB* and *PLCG1*. The authors believe the resulting aberrant signaling in the pathway is likely the driver for AS [128].

There is no cure for AS, though if the disease is diagnosed soon after inception, surgical intervention and chemotherapy may prolong patient survival [129]. Nevertheless, the nature of AS growth is somewhat deceptive, and the clinical signs are not visible until it is too late. Canine HSA is being recruited as a model to better understand the pathology of the AS disease [118, 130]. The information gathered from these research efforts could potentially benefit both species with new therapeutic strategies.

## 2.4 Glioblastoma

Gliomas are dominant forms of brain tumor, second only in incidence after meningioma, that arise from the neural stem or progenitor cells [131]. Gliomas encompass a group of cancers, which include astrocytic tumors, ependymomas, oligodendrogliomas, and mixed gliomas [131]. CNS tumors are graded on a scale of I-IV: at one end of the spectrum are the grade I pilocytic astrocytomas, which are slow-growing and benign tumors [132], and at the other the maximum grade IV glioblastoma (GBM), the most common as well as the most malignant form of glioma [133]. In 2016, the World Health Organization (WHO), based on both histology measurements and molecular factors, has restructured the classification of tumors that arise in the CNS [134]. Two forms of GBM are demarcated, based on the status of mutations found in the isocitrate dehydrogenase *IDH1* and *IDH2* genes [135]. The primary tumor, wild-type for *IDH*, is exceptionally aggressive, occurs *de novo*, and accounts for ~90% of GBMs [136]. It is often diagnosed in the elderly (>55 years), and the affected patients have a median survival period of about 9.9 months [137]. The secondary tumor, which often progresses from low-grade diffuse astrocytoma or anaplastic astrocytoma, carries mutations in the *IDH* gene [136]. It has relatively less necrosis when compared to the primary form and also has a better prognosis, with median survival being around 24 months [137]. In the WHO classification, there are reports of sporadic cases where the tumors are grade IV GBMs, but their *IDH* mutation status cannot be ascertained and are therefore labeled as ‘not otherwise specified’ (NOS; [134]).

As of today, the etiology of GBM remains unknown. There are few studies that have focused on the probable inherited susceptibility to the disease. For

example, Melin and colleagues [138], in a meta-analysis of three GWAS studies, identified nine risk loci (four of which had been previously reported) linked with GBM. These include 5p15.33, 7p11.2, 9p21.3, and 20q13.33; however, they have not ascertained candidates for causality among the associated loci. Loss of heterozygosity in chromosome 10, observed in up to 80% of the primary tumors, is the most common genetic alteration in GBM [139-142]. Another frequent alteration is the co-deletion of the arms of 1p/19q seen in GBM tumors that tend to have better clinical outcomes [143-146]. In a somatic mutational analysis of GBM with more than 500 tumor samples using NGS approaches, Brennan *et al.* [147] distinguished recurrent point mutations in known GBM genes that included the tumor suppressors *TP53*, *PTEN* and *RB1*, and the oncogene *PIK3CA*; in addition, they also found frequent changes in ~60 novel genes that may have roles in the disease progression. New structural rearrangements of signature receptors were also uncovered in the study.

Inter- and intra-tumor heterogeneity are known to be prominent in GBM [148, 149]. For instance, in an investigation involving 11 patients, genomic analyses of multiple tumor sections of individual tumors showed variable changes reflecting intra-tumor heterogeneity [150]. These changes were also shown to correspond to different molecular subtypes within the same tumor. In an RNA-seq study [151], expression profile analysis of 430 cells from five primary GBM tumor samples showed that genes related to proliferation and oncogenic signaling, among others, were differentially expressed in the cells and the tissues they originated from, indicative of intra-tumor diversity.

Surgical resection, followed by radiotherapy and chemotherapy with temozolomide, form part of the standard care for GBM [152]. In most cases, though, the tumor reappears within a short time, spreads to other regions of the brain, and ultimately leads to the death of the individual. A recent investigational effort that focuses on tumor-localized gene therapy [153] indicates better survival for the patients in the study. Vaccination to stimulate immunosurveillance against glioma cells is being actively tested in several clinical trials [154]; however, negative results from some of earlier phases of the trials deem that vaccination can only be part of a multi-modal treatment approach. Another innovative approach involves an oncolytic (virus)-based therapy with the drug PVSRIPO (Polio/Rhinovirus Recombinant) that is currently being tested on  $n = 61$  patients with recurrent GBM (<https://clinicaltrials.gov/ct2/show/study/NCT01491893>). While studies like these show promise, it remains to be seen how effectual they will prove to be when applied to a larger population.

A footnote to include in the present section, despite the subject not being part of the current investigations, is with reference to glioma in dogs. Spontaneous gliomas in dogs account for about 35% of all tumors affecting the CNS [155]. Though epidemiological evidence for the prevalence among breeds is lacking, based on necropsy investigations in 21 breeds [156], there was evidence for glial neoplasms being common in brachycephalic dogs. In 2016,

Truvé *et al.* [157] from our lab were involved in a GWAS study, which identified three genes that have associations with glioma; these include *CAMKK2*, *P2RX7*, and *DENR*. The authors surmise that because of the histological and genetic similarity that exists between human and canine glioma, the above associations may be relevant for the study of the corresponding human disease.

## 2.5 In summary

The cancers focused on in this thesis are all characterized by significant complexity, encompassing cytopathological, transcriptional, and genomic levels. Also, dealing with their inter-tumoral and intra-tumoral tumor heterogeneity, to how cells become cancerous and acquire metastatic capabilities, are all factors that make these cancers challenging to comprehend as a researcher, and to clinically combat as a patient.

Sequencing approaches adopted in the current studies potentially offer avenues to uncover some of these cancers' molecular mechanisms and gain a better understanding of their pathobiology.

## 3 The current studies

### 3.1 Aims of the thesis

Somatic changes resulting from errors during DNA replication and exposure to endogenous or exogenous mutagens accrue in the genome during the process of cell division. Some of these aberrant modifications can provide cells with a selective growth advantage that can result in the disease of cancer. These selected alterations – often termed driver mutations – can occur either in coding or non-coding, or in both regions of the genome, and trigger carcinogenic processes. The overall aims of the thesis, therefore, were to characterize the somatic mutational landscape across three types of malignant tumors prevalent in dogs and humans.

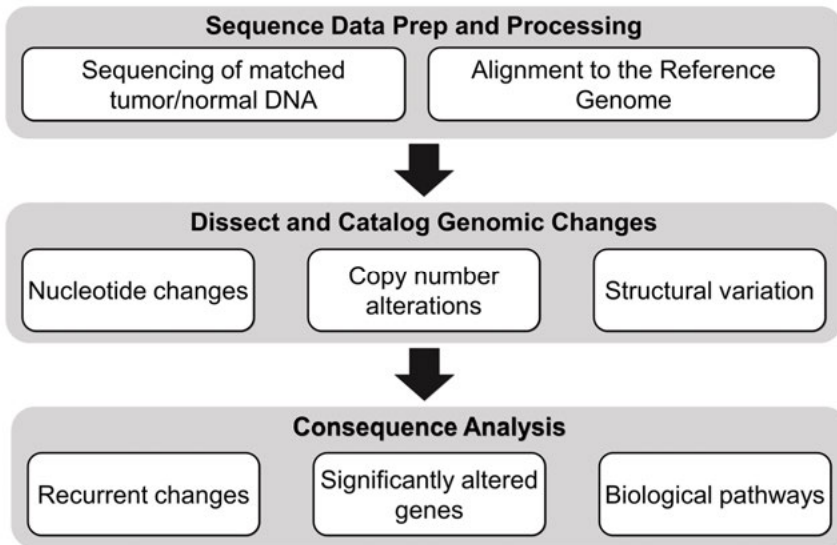
Explicitly, the aims are:

- I. To utilize whole-exome sequencing methodologies to distinguish coding mutational changes that may have roles in the initiation and progression of osteosarcoma tumors, across three dog breeds with differential susceptibility to the disease.
- II. To delineate coding changes present in canine hemangiosarcoma using whole-exome sequencing, and perform a comparative analysis with the somatic modifications observed in the analogous human tumor of Angiosarcoma to converge on driver alterations that may be pervasive across species.
- III. Using whole-genome sequencing to identify and prioritize coding and non-coding genomic changes that may have functional roles in the human tumor of glioblastoma.

## 4 Remarks on methods implemented in the current studies

### 4.1 Overview of cancer genome analysis

The sequential workflow for a standard cancer genome analysis employed for the discovery of somatic variants that explain (some of) the cancer biology across the studies in the thesis is shown in **Figure 3**.



**Figure 3.** Standard pipeline for cancer genome analysis with matched tumor-normal sequencing reads. Nucleotide changes can be either point or indel mutations. Copy number alterations can be either focal- and/or large-scale-, amplifications, or deletions.

The first step, irrespective of specific cancers, involves the biological sample collection after ensuring that ethical certifications and informed consent are in place. The samples in the cohort are then sequenced to appropriate targets (whole-exome or whole-genome) and depths of coverage set for the study design. The second step involves performing sequencing and aligning the sequenced reads to a reference genome assembly, to identify the genomic coordinates to which each read fragment maps. With the aligned reads, the next

stage is variant calling with appropriate protocols. The purpose here is to discern the differences that may exist between the sequenced reads and the reference genome that exist only in the tumor. These differences can be small or large, from a single nucleotide change to several thousand base pairs changes. The sequences could also potentially be rearranged with duplications, inversions, or complex rearrangements.

It should be emphasized that the somatic variant calling in cancer genomes is dissimilar to the germline detection of variants, and the algorithms used for the latter's delineation are not suitable for the former. For germline calling, the assumptions are that the genomes are diploid, that the alternate allele (the mutant base) frequency for a heterozygous call is 50%, and that an alternate homozygous call is 100%. On the other hand, the somatic variant callers do not have any supposition about the ploidy levels and have to be designed to call variants at very low allele fractions. They also need to contend with normal (healthy cells) admixture in the tumor sample, or the variants could come from only a fraction of the tumor cells (subclonal variants). Additionally, copy number or ploidy changes in the cancer genome could introduce artifacts in the calling that have to be appropriately identified and removed.

Most somatic variant algorithms are devised to analyze matched tumor-normal samples (where the sequenced DNA for these come from the same individual), and to look for differences between them and also for dissimilarities between the tumor reads and a reference genome. The calls that are unique to the tumors are then marked as somatic variants, and the consequences of these variants are then assessed. Though most of the variations discerned are likely to be benign, if they fall in a protein-coding region of the genome, it could potentially alter the function of the protein, which could be deleterious. If the variants happen to be in critical non-coding regions of the genome – for example, in the regulatory regions – it can considerably affect the regulation of genes and pathways involved in cancer [158]. Recurrent variants (somatic calls identified in more than one individual tumor) are an important indication of driver status in cancer genomes [159]. The mutational ‘hotspots’ in which they occur could potentially be directly involved in the carcinogenic process, and the identification of these positions can likely serve as biomarkers [160]. Lastly, in the cancer genome analysis, pathway enrichment offers a comprehensive method for connecting mutated genes with known signaling pathways and regulatory networks to understand the mechanisms and biological functions they may be involved in [161, 162].

## 4.2 Whole-exome and whole-genome sequencing of cancer

Exome sequencing or whole-exome sequencing centers on the protein-coding sequences, which constitute approximately 1% of the entire genome [163]. This is accomplished through an enrichment procedure, where DNA baits are

used to hybridize with only the protein-coding parts, separating these regions from the non-coding portion of the genome [164]. WES has been effectively implemented in several disease-gene detection as well as diagnostic projects [165, 166]. In cancer studies, with WES, it is possible to discover driver mutations in genes (ones that are responsible for either disease initiation or progression) without any prior knowledge of the actual genes [167-169]. However, there are a few drawbacks to WES. For instance, in the detection of copy number alterations that may span beyond the coding/captured part of the genome, WES introduces biases that make their detection difficult [170]. Also, there are issues in detecting structural variants (SV) at base-pair resolution where the junction breakpoints cannot be clearly defined [171].

Whole-genome sequencing of cancer samples entails sequencing the vast majority of the genome. With WGS, in addition to being able to probe the coding sequences, it can, unlike WES, also be used to detect multiple classes of alterations – copy number variations (CNV), rearrangements, and other SVs – with more confidence and certainty [171]. Also, it is possible that the non-coding genome contains pathogenic variants [172], whose detection is likely to be missed when using targeted approaches like WES. These include point or indel mutations in cis-regulatory regions, which can create or disrupt transcription factor-binding sites [158, 173], or in long noncoding RNAs (lncRNAs; [174]) which are aberrantly expressed in various cancers [175-177]. WGS circumvents most concerns that WES has; nevertheless, there are still a few disadvantages. Though CNV and SV calling is efficient with WGS, somatic indel calling shows a significant discrepancy, as discerned in the concordance analysis using various algorithms [178]. Another challenge for WGS is in assembling the requisite computing structure for storage and analyses in a cost-effective manner [179].

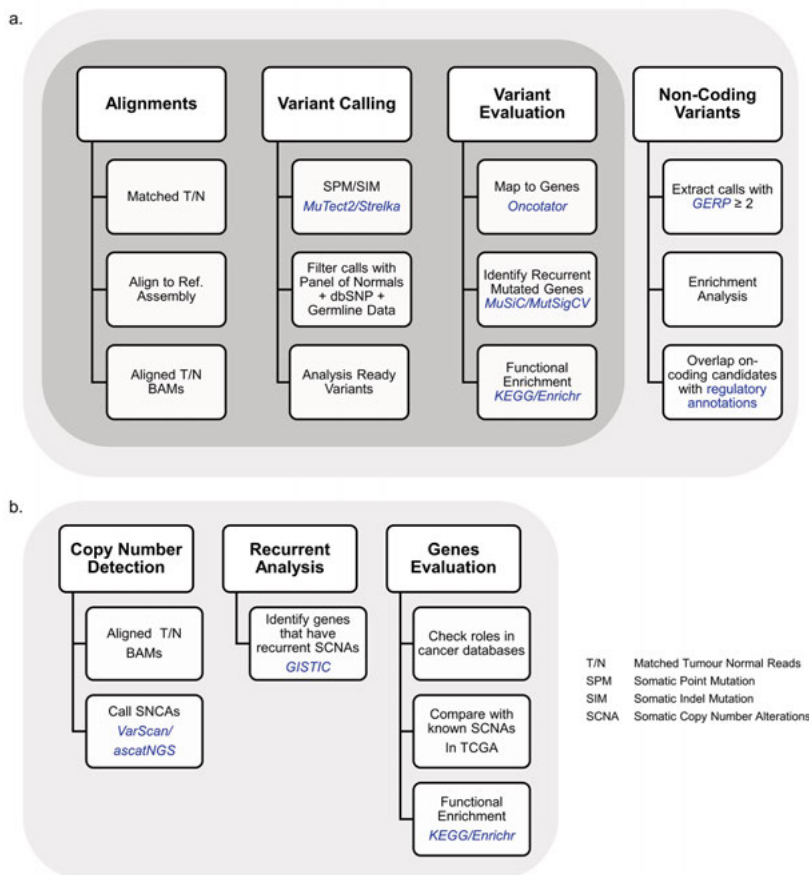
The consideration regarding the use of WGS or WES methods hinges on whether a complete or a more-targeted approach to characterization of the cancer genome is warranted for a given study. It is estimated that WES sequencing per patient is about a third cheaper than what it is with WGS [180], and, consequently, the associated overheads also may play a decisive role in the choice of technology adopted for a given study. Ultimately, though both WGS and WES offer prospects for the discovery of carcinogenic alterations, their scopes are vastly different – from the regions of the genome scanned to the turnaround time for delivery of analysis-ready results, and budgetary constraints; these are all factors that have led to the variations in their particular implementation.

The WES or WGS methods implemented for each of the studies in the thesis are detailed in the attached published or submitted manuscripts.



### 4.3 Bioinformatics analyses of cancer genomes in the current studies

Standard bioinformatics pipelines for cancer genome analysis were employed to elucidate sources of variation that are present in the diseased tumors. A typical protocol is shown in **Figure 4**. Quality-controlled passing sequencing reads were aligned against the dog [181] and human reference genomes of CanFam3.1 and hg19 [31], using the tool BWA [182]. After a multi-step refinement of the resulting alignments, somatic point mutations (SPM) and somatic indel mutations (SIM) were called using the algorithms of MuTect2 [183] and Strelka [184]. To reduce miscalled germline events and sequencing artifacts, the raw calls are subjected to successive filtration steps. This involves the removal of polymorphism calls that have been marked in the panel-of-normals, or in the publicly available germline databases like dbSNP [185] or SweGen [186]. The final SPM and SIM thus obtained were then annotated with gene information using the tools of SnpEff v4.2 [187] for the dog cancers of OSA and HSA, and Oncotator [188] for the human GBM. In the coding part of the genome, the tools of MuSiC [189] and MutSigCV [190] were employed to distinguish putative driver mutations from passengers (mutations that do not have an impact on cancer progress). Ensembl's Variant Effect Predictor (VEP; [191]) was used to prioritize genes with high-impact mutations.



**Figure 4.** The bioinformatics pipeline adopted for the cancer genome analysis. 4a. Workflow for calling somatic point and indel mutations. The dark grey inset box is valid for WES matched reads. The light grey encompasses methods applicable to WGS. 4b. Workflow for copy number detection. The tools used are shown in blue.

To delineate non-coding alterations that may have a functional impact (in the cancer of GBM, where the entire cancer genomes was sequenced), we looked at likely functional non-coding variants around selected GBM genes based on the 33-mammals alignment constraint scores as defined by GERP++ [91]. The non-coding constraint mutations (NCCMs; variants found in regions of sequence conservation in the neighborhood of the above genes) were annotated with regulatory information downloaded from either the UCSC genome browser [192] or ENCODE portal [89]. These included data from tracks of transcription factor binding sites (TFBS), methylation markers, regulatory markers, transcription start sites, and enhancer information, among others. Mutations that were likely to affect non-coding constraint regulatory regions were further studied for transcription factor binding affinity, using the tool

‘TRanscription Factor Affinity Prediction’ (sTRAP program for detecting differences in binding between two sequences [193]).

Mutational signatures, ‘imprints’ of somatic mutations hidden in the cancer genome, were detected using the Bayesian variant of the nonnegative matrix factorization (NMF) algorithm [194] for the dog cancers, and using the online utility ‘MUTation AnaLyIS toolKit’ (Mutalisk; [195]) for the GBM study. The results obtained were compared with the 30 COSMIC signatures ([196]; <http://cancer.sanger.ac.uk/cosmic/signatures>).

Somatic copy number alterations in the samples were detected using the tools of VarScan2 [197] in the dog cancers of OSA and HSA, and ascatNGS [198] in the human GBM dataset. For all the copy number alterations discerned, GISTIC v2.0 [199], a tool that evaluates the frequency and fold-changes of the copy number alterations, was applied to prioritize regions that may drive cancer development.

Pathway analysis, to identify a shared network the significantly mutated genes may be involved in, was performed using multiple methods across the projects. These include the online utilities of DAVID [200], Enrichr [201], and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [202].

#### 4.4 Statistical analysis employed

Descriptive statistics was used to tabulate the mutational and copy number alteration frequencies for all of the studies.

In Study I, Spearman’s correlation was run to determine the relationship between age and total counts of somatic mutations found across the breeds. The tools of MuSiC and MutSigCV assessed the significantly mutated genes, and for both, a false discovery rate (FDR) threshold of 0.1 was applied. In the identification of putative drivers with VEP’s high-impact mutations among the non-silent genes, the Wilcoxon rank sum test, with multiple testing corrections for the number of genes tested, was implemented. Functional annotation charts were created for KEGG using the default options, using the Benjamini–Hochberg method to control the FDR.

In Study II, methods for MuSiC and pathway enrichment analysis with KEGG were run as in Study I.

In Study III, statistical tests for enrichment of non-coding constraint mutations around GBM genes as well as other protein-coding genes were performed with unpaired *t*-test, using the R statistical framework (R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>).

## 5 Summary of current studies

### 5.1 *SETD2* Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma (Paper I)

#### 5.1.1 Background

To characterize somatic alterations in canine OSA tumors, we performed WES of matched tumor-normal (T/N) pairs in three breeds GR, GH, and RW, all predisposed to osteosarcoma but with different lifetime risks. The goals here were to discover putative candidate cancer driver genes, to evaluate if OSA in dogs shares mutated genes with OSA in humans, and, lastly, to detect potential differences among breeds.

#### 5.1.2 Results

Sixty-six pairs of samples (22 GR, 23 GH, and 21 RW) passed the quality thresholds, with a median sequence depth of 136x (range, 90-176) for the tumor and 95x (range, 31-166) for the normal samples. Across all three breeds, VarScan2 identified a large number of SCNAs. For both amplification and deletion SCNA classes, the GRs had more of the genome altered than either the RWs or GHs. GISTIC v2.0 identified 19 large-scale and 67 focal recurrent SCNA events. The chromosomal regions corresponding to recurrent focal SCNAs contained numerous cancer-related genes, including *TP53*, *CDKN2A/B*, *HRAS*, and *PTEN*.

Analysis of the exome sequence data using MuTect2 revealed a total of 7,900 SPMs and 1,197 SIMs across all individuals. The average numbers of SPMs and SIMs were different among the breeds – the GRs had a more extensive range of mutations, showing a higher median of 133 SPMs, compared with 113 SPMs in RWs and 97 SPMs in GHs. However, when the SPMs and SIMs were mapped to gene information with SnpEff, followed by enrichment analysis with MuSiC, 11 genes were found to be significantly mutated across all three breeds. The top SMGs across all samples were, in order, *TP53*, *SETD2*, and *TANGO*. *TP53*, tumor protein 53, had mutations in roughly 60% of all samples, and *SETD2* was mutated in 21% of samples, with a variety of

non-silent mutation types (frameshift, nonsense, splice, and missense mutations). It is worth noting that the *SETD2* gene has not previously been linked to osteosarcoma.

The mutational signature analysis identified three distinct mutational profiles, including COSMIC 1, the aging-associated signature, and COSMIC 17, a motif for which the etiology has not been established in humans. The third signature, characterized by T>G transversions, is novel. Although all three signatures were found in each of the breed groups, the contribution of each signature per sample per breed indicates that there is a difference in the representation of signatures. GHs and RWs have a higher rate of COSMIC 1, whereas the novel signature is more prevalent in the GRs.

It is known that in humans, germline mutations in the genes of *GRM4*, *WRN*, and *BLM* (the last two are associated with specific familial syndromes) increase the risk for OSA. In the dogs in our study, we find candidate germline mutations in the above susceptibility genes; however, there is variability in the prevalence of these mutations across the breeds, *i.e.*, specific genes are seen to be mutated exclusively in one or two breeds and show no changes in the other(s).

### 5.1.3 Discussion

The somatic mutation landscape in the coding genome in dog breeds predisposed to OSA was evaluated using WES approaches. Our findings reiterate and extend what has been reported for the corresponding human disease. These include widespread recurrent copy number aberrations and fewer point mutations in all the three breeds studied. The above SCNAs encompass several tumor suppressors and oncogenes, some of them with known roles in OSA and other genes are shown here for the first time in the disease. *TP53* is the most frequently altered gene (with >80% of the dogs having either mutations or SCNAs), again reflecting human OSA patterns. *SETD2*, a histone methyltransferase (HMT) and an epigenetic regulator, though it has no recognized ties to OSA cancers yet, has known tumor-suppressor activity across several human cancers [203-205]. Here, in the canine tumors, we find that the 16/17 non-silent changes in *SETD2* (as predicted by VEP) are high-impact mutations, *i.e.*, mutations that are likely to have a disruptive effect on the gene and lead to its loss of function.

Additionally, support for the finding that histone modification enzymes having roles in OSA was further strengthened when we uncovered that 36% of the dogs in the study have mutations in either *SETD2* or other members of this family of genes that include histone acetyltransferases (HATs) and histone deacetyltransferases (HDACs). Histone modifiers, when disrupted, are known to cause cancerous alterations and promote tumor proliferation [206]. Inhibitors developed for these druggable ‘targets’ are in clinical trials for several cancers [207] and so can conceivably also be used for OSA.

Mutational signature analyses used for discerning patterns of somatic changes confirm the existence of three distinct signatures, including a novel signature, across all breeds. Each signature, though, has different distributions of signatures per breed. COSMIC 1 is widely seen in both the GH and RW samples and is also the dominant signature in human OSA [208], and the novel signature is seen almost exclusively in the GRs. Breed differences were also detected in germline mutations among candidate protein-altering or predisposing genes, indicating a probable role for the genetic background in tumor initiation. In summary, the similar spectrum of SCNAs and the mutation burden seen cross-species works not only to inform the OSA biology but also to highlight the usefulness of the canine disease models for enhancing our understanding of the corresponding human disease.

## 5.2 Comparative Genomics Reveals Shared Mutational Landscape in Canine Hemangiosarcoma and Human Angiosarcoma (Paper II)

### 5.2.1 Background

To discern the somatic mutational spectrum of canine HSA, we implemented WES of primary tumors and matched healthy tissues from an HSA-susceptible breed, the golden retriever. In addition, we wanted to assess if the clinical correspondence of the disease to human AS also had comparable genetic etiology or genomic underpinnings.

### 5.2.2 Results

Matched T/N WES reads from golden retrievers ( $n = 47$ ) were aligned against the dog reference genome to an average sequencing depth 78X for the tumors and 63X for the normal samples. A standard pipeline for calling somatic variants was implemented with the MuTect2 algorithm and followed by functional annotation of the resultant filtered calls using SnpEff. Significantly mutated genes were then identified using the tool MuSiC. Seven SMGs were discerned across the dataset, including mutations in the tumor suppressor *TP53* (28/47 tumors, ~60%), in the oncogene *PIK3CA* (14/47 tumors, ~30%), and in *PIK3R1* (a regulatory subunit of *PIK3CA*, 4/47 tumors, ~9%). Mutations at amino acid position 1047, encoded by the gene *PIK3CA*, is a known hotspot for mutation across several human cancers [209]. We find that 10 of the 14 tumors that have mutations in the *PIK3CA* gene gave rise to either H1047L or H1047R changes. Though not frequently mutated, there are other genes from the *PI3K* gene family that have non-silent mutations in the cohort, including *PIK3CB*, *PIK3C2G*, *PIK3C3*, *PIK3R1*, and *PIK3R5*. We then compared the somatic mutations observed in our cohort with the reported changes in ‘The

Angiosarcoma Project' (a patient-led effort that performs sequencing of AS tumors; <https://ascproject.org/>) and found that the frequently mutated genes here include the canine top SMG hit *TP53*. Analysis of mutational signatures for the cohort revealed that the dominant signature, COSMIC 1, is in accord with what was seen as the main profile for the human AS [210].

KEGG pathway enrichment analysis [202], with datasets from canine HSA and two subtypes of human AS (defined based on the tumor location), showed there were many shared pathways between the two diseases. The angiogenesis pathway genes of Phospholipase C gamma 1, *PLCG1*, and Protein Tyrosine Phosphatase Receptor Type B, *PTPRB*, are known to have recurrent mutations in human AS (mentioned elsewhere/noted in the background section). Neither of these genes, though, has recurrent alterations in canine HSA, but seven tumors were found to harbor mutations in other PLC genes; likewise, eight cases had mutations in the PTP family of genes.

Lastly, copy number analysis shows the alterations recur in many cancer-related genes, including the tumor suppressor of *CDNK2A/B* (deleted in 22% of the samples) and the *MYC* oncogene (gains in 9% of the samples). Comparative analysis for SCNAs between the human AS data and canine oligonucleotide array comparative genomic hybridization (oaCGH) data showed copy number aberrations in known cancer genes, including *KDR* and *AXINI*, for both species. Interestingly, the gene *KDR* was also found to be significantly mutated in human AS patients and is observed to be mutated only once in the canine HSA cohort with a predicted benign or tolerated effect.

### 5.2.3 Discussion

The genetically homogenous population of golden retrievers was used to delineate the molecular profiles for canine hemangiosarcoma. Our investigations reveal there are both genetic similarities as well as differences that exist between the dog HSA and the human AS. *TP53* is the only gene found to be significantly mutated in both diseases. In the canine dataset, the majority of the mutations for *TP53* occur in the DNA-binding domain, likely leading to loss of function for the encoding protein. The second SMG, *PIK3CA*, which is known to exhibit oncogenic activity across several cancers [211], is found only in a subset of human AS (only in the breast tumors). There are also several other genes in the PI3K pathway that were commonly mutated in our cohort. The PI3K pathway is an essential cellular pathway, and one of the most frequently altered pathways in cancer, playing an essential role in signal transduction, leading to cell proliferation, survival, differentiation, and regulation of metabolism and immunity [212, 213]. Another potentially important difference between the two species is that, while copy number gains in *KDR* are typical in both species, somatic mutations in this gene were seen in over 20% of human tumors but in only one canine tumor. As the *KDR* receptor is upstream of the PI3K pathway, it is possible that mutations in either may lead to

a similar phenotype. Tumors in both dogs and humans were enriched for mutations in protein tyrosine kinases, which are essential regulators of cellular growth and division signals and are commonly mutated in cancers. There were also recurrent mutations in the protein tyrosine phosphatase gene family in both species. Tyrosine kinase inhibitors have been used for angiosarcoma in the clinic and have also shown promise against hemangiosarcoma *in vitro* [214], but so far have been less favorable in the clinic [215]. Investigations of the interaction between the many affected pathways, therefore, may help to determine the possibility of designing a combination therapy for targeting them. Summing up, our characterization of the biological foundations of canine HSA reveals many similarities but also some important differences with human AS; this knowledge hence can be used to better understand the pathogenesis of both diseases and come up with therapeutic strategies that will serve both species.

### 5.3 Whole genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes (Paper III)

#### 5.3.1 Background

To systematically identify the genomic terrain encompassing both the coding and non-coding cancer genome of the brain tumor glioblastoma, we WGS matched tumor samples from  $n = 39$  (human) patients who did not carry a mutation in the *IDH1/2* gene (*i.e.*, were *IDH1<sup>wt</sup>*). The principal objectives for the study were: first, to verify if the coding landscape reiterated observations from previous investigations into the disease; and next, and more importantly, to explore and prioritize somatic changes in non-coding sequences, especially in regulatory regions that may have possible pathogenic consequences in GBM.

#### 5.3.2 Results

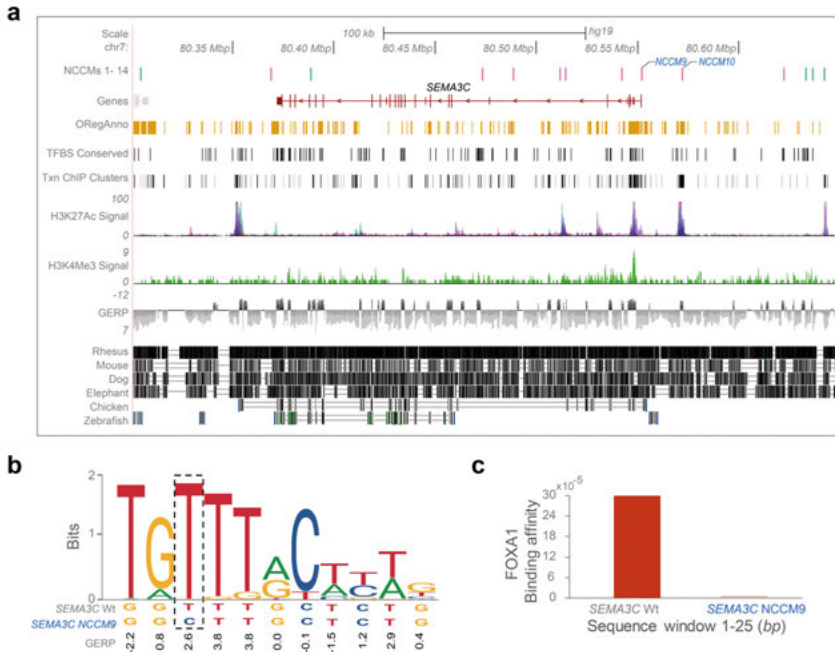
The matched tumor-normal DNA from the cohort of 39 patients was sequenced using Illumina whole-genome sequencing methodologies. Alignment of the reads to the human reference assembly yielded depths of coverage of median 75x (range: 64-89) for the tumors and 38x (range: 30-66) for the matched normal. SCNAs were detected using the tool *ascatNGS*, and the patterns were found to highly similar to what was seen in the ‘The Cancer Genome Atlas’ (TCGA) cohort [216, 217]. As in the previous two studies, *MuTect2* was next employed to discern somatic point and indel mutations. Here,



however, to achieve a highly reliable dataset, we decided to intersect the above mutations with Strelka's call set, and only those SPMs and SIMs that were concordant between the two callers were used for downstream analysis. Annotating the variants with Oncotator, followed by the analysis of the protein-coding genes with MutSigCV, showed *TP53*, *PTEN*, and *EGFR* as the three SMGs. Also, we adopted a frequency-based approach for delineating more coding genes that may potentially have driver alterations. Accordingly, we found nine other frequently mutated genes (FMG) that harbored non-silent mutations in  $\geq 10\%$  of the study cohort. The 12 genes from the SMGs/FMGs list overlap with the top 20 genes seen to be mutated in the TCGA-GBM dataset [218].

The majority of SPMs and SIMs in our cohort were in the non-coding regions of the genome. Most of them are likely to be passengers; nonetheless, a fraction of these non-coding changes which are associated with regulatory elements such as promoters, UTRs, splice signals, enhancers, lncRNAs, and TFBSs, among others, can be expected to have roles in carcinogenesis [172]. Frequent alterations in the promoter of the telomerase reverse transcriptase, *TERT* gene, have been described across several cancers [219, 220], including glioma [221]. In our cohort, we find that  $>75\%$  of samples have mutually exclusive mutations at two positions previously reported for the promoter [221]. We decided to investigate further mutations in the regulatory sequences of genes known to have roles in GBM, and we focused only on variants occurring in evolutionary constrained sequences in the vicinity of these key genes. (The list of 78 key genes was compiled by combining the SweGBM-1 SMG/FMG and TCGA-GBM SMG gene sets). By applying GERP constraint scores, we find a significant enrichment of NCCMs associated with the 78 key genes versus all other protein-coding genes.

Twenty-six of these 78 genes had  $>1.0$  NCCM per 100 kbp. These NCCMs were annotated with functional datasets, such as TFBS, acetylation, and histone markers, among others. Mutations in these functional sites can potentially have an impact on GBM. For instance, the GBM gene *SEMA3C*, with a total of 14 NCCMs, intersects several regulatory annotations (**Figure 5a**). NCCM 9, associated with this gene in the study cohort, had a mutation in its promoter that was predicted by sTRAP to disrupt the binding of the transcription factor *FOXA1* (**Figure 5b and 5c**).



**Figure 5.** UCSC genome browser view of *SEMA3C*, the top key GBM gene with the highest rate of non-coding constraint mutations. 5a. The NCCMs seen in both introns as well as the in the flanking regions lie regions of genome that are well conserved across mammals and have at least one regulatory annotation. 5b. The sequence logo (MA0148.1) of the FOXA1 TFBS shows that the *SEMA3C* NCCM9 mutation affects a highly conserved nucleotide that could abate the binding in the mutated site compared to the wildtype. 5c. The affinity profiles, for the same mutated sequence, shows a decreased affinity for the FOXA1 transcription factor, compared with the wild-type sequence.

In another key gene, *DYNC1H1*, 9/20 NCCMs overlap with the activating epigenetic marker H3K27me3 (tri-methylation on lysine 4 of histone) that is often found in promoter regions and is closely associated with transcriptionally active genes. Among the other protein-coding genes, a total of 1,776 genes had a frequency of >1 NCCMs/100 kbp, of which 43 had  $\geq 3.0$  NCCMs per 100 kbp sequence, indicating that there are genes and associated regulatory regions that possibly have roles in GBM. These include Distal-Less Homeobox 5, *DLX5*, which has been shown to affect glioma cell motility via the PAX6/DLX5-WNT5A axis [222]. On chromosome 14, four genes—*SLC25A21*, *MIPOL1*, *FOXA1* and *TTC6*—had over 55 shared NCCMs that not only are in regions of high conservation but also overlap multiple transcription factor binding annotations. Analysis with the utility sTRAP showed significant differences in the binding affinity in the mutated sequences versus the corresponding wildtype sequence.

Examination of the mutational signatures for the cohort with Mutalisk revealed two dominant profiles – COSMIC 1 and COSMIC 5 – which is also in accord with previously detected signatures in GBM [196]. Interestingly, when the variants were split into coding and non-coding alterations, the former showed predominance for the aging signature, whereas the non-coding variants showed near equal distribution between the COSMIC 1 and COSMIC 5 signatures.

### 5.3.3 Discussion

Although somatic alterations in the genome are responsible for the disease of cancer, most studies in GBM, as well as in other cancers thus far, have stressed the discovery of coding driver mutations. The non-coding sequences which account for >98% of the genome are largely left unexplored. In this study, we set out to distinguish the non-coding mutations that arise in GBM tumors that may have roles in carcinogenesis. Though our cohort of  $n = 39$  is relatively small, the protein-coding alterations, SCNAs, and mutational signatures found here mirrored what was previously seen in the TCGA-GBM dataset. This, therefore, lends credibility to our cohort and the findings derived from it.

Mutations within the functional, non-coding regions of the genome can alter gene expression, splicing, expression of non-coding transcripts, and the epigenetic state [172]. However, the question of how to prioritize the mutations in the non-coding regions into drivers and passengers is a challenge. Evolutionary conservation provides a practical way of identifying which specific positions are likely to be important for genome function [223]. We, therefore, used the concept of mammalian constraint (as defined by GERP++) to hone in on putative non-coding mutation candidates around 78 key GBM genes. We identified highly significant enrichment of NCCMs in the neighborhood of the key GBM genes, with 26 genes found being enriched for NCCMs (>1 NCCMs/100 kbp). The key genes with the highest frequency of NCCMs were *SEMA3C*, *DYNCH1*, *CNTNAP2*, and *LRFN5*. While the NCCM frequencies in key GBM genes were skewed to the right, compared to all OPG, the latter category still contained genes with high NCCM frequency, including a total of 43 genes with >3 NCCMs/100 kbp. To assign candidate functions to the NCCMs, we used publicly available sources of genome annotations in addition to the evolutionary constraint and found that most NCCMs had functional annotations, suggesting that mutations in them could potentially be drivers. We also observed that 91% of NCCMs in the top genes (26 key GBM genes and 43 from the OPG set) had a variant allele fraction of  $\geq 10\%$ , again supporting their ability to affect the tumor initiation and/or progression. (Variant Allele Fraction or VAF is the fraction of reads overlapping a genomic coordinate that supports the alternate allele).

When assessing the potential biological importance of the analyzed variants, *SEMA3C* was conspicuous, both due to its large number of NCCMs and

due to indications in the current literature that overexpression of the gene is linked to poor prognosis in several cancer types, including prostate cancer [224] and GBM [225]. The *SEMA3C* gene has been shown to be regulated by several transcription factors, including FOXA1, GATA2, and GATA6. Mutations occurring in TFBSs can disrupt the binding of transcription factors (TFs) and lead to gene dysregulation/aberrant gene expression, which can trigger carcinogenesis [226]. We also observe that multiple NCCMs overlap binding sites for the GATA family of transcription factors. Given that GATA factors are known to coordinate cell survival, cellular maturation, and proliferation arrest [227], this family of genes has been anticipated to have a role in human cancers [228]. Variants for *SEMA3C*, *DYNC111*, and *CDH18* lie in conserved TFBS of the GATA family. GATA2 has been directly implicated in the promotion of glioma through the EGFR/ERK/Elk-1 pathway [229], further indicating its potential to forward tumor development in GBM.

The discovery of non-coding driver mutations in GBM is still a nascent area in the field of cancer genomics. In this study, we show that it is possible to use the concept of evolutionary constraints in conjunction with relevant genomic annotations to identify candidate mutations. We hope that the conclusions from this study will provide a basis for further analysis of pathogenic non-coding variants, not only in the cancer of GBM but also potentially in other cancers too.

## 6 Concluding remarks and future directions

### 6.1 Overview

The objectives of the work in the current thesis were to increase our understanding of three diverse cancers: osteosarcoma, hemangiosarcoma, and glioblastoma. The goal was achieved by the demarcation of somatic alterations that are imprinted in the genomes of the above diseases. The highlights of each study are summarized again below.

In the first study of osteosarcoma in three dog breeds with differing susceptibility to the disease, it was shown that the mutational landscape at the nucleotide level, though discordant, with more private mutations per breed, all converge on the same (putative) driver genes. These include the genes of *TP53* and *SETD2*. The tumor protein p53 encoded by the gene *TP53* is not only responsible for maintaining the genome integrity of an organism, but in response to an array of cellular stimuli/stresses, also promotes cell cycle arrest, apoptosis, and DNA repair, among its several other functions [230]. Mutations to the gene can be extremely damaging, leading to its loss of tumor-suppressive function and resulting in carcinogenesis, as has been observed in a multitude of cancers [231, 232]. *SETD2*, an epigenetic modifier, has, in recent years, also been implicated as a tumor suppressor in various types of cancers [205, 233, 234]. Although our study was the first to expressly link *SETD2* with OSA, others have since corroborated its role in the disease [235, 236]. Gardner *et al.*, in their investigation across multiple dog breeds, elaborated on the biological consequences of *SETD2* alterations, which they found to be recurrently mutated in >40% of their tumors [236]. Also, support for epigenetic modifications was reinforced by their discovery that 67% of the samples showed aberrations in epigenetic and chromatin-modifying genes. Leveraging the knowledge gained about epigenetic mechanisms in OSA can, therefore, be used to serve as a valuable biomarker, and also to explore the efficacy of epigenetic therapy to inhibit the disease.

In the second study centering on hemangiosarcoma, the coding mutational profile was delineated in a single-breed study, wherein we found driver mutations in the genes of *TP53* and the oncogene *PIK3CA*. The dog HSA is the histopathological equivalent of the human disease AS, with both showing the same aggressive clinical course. Our investigations reveal that the genomic/coding bases also share many similarities, from similar mutated genes

to copy number alterations and biological pathways. Human AS is an uncommon cancer (accounting for only 0.01 of all cancers; [237]), and therefore any genomic investigations that can be undertaken are likely to be under-powered. However, given that dog HSA shows similar pathobiology with AS and also the fact that it is a vastly more prevalent cancer in canines, we can use them as models to study the human disease with relative ease (and power). Our study is a step in this direction, where our findings show that prospective therapeutics designed based on the genomic targets discovered in dogs may also possibly inform the human treatment of the disease.

In the last study, genome-wide characterization of the somatic alterations in the cancer of glioblastoma was embarked upon. At the outset, we find the coding alterations recapitulate what has been previously seen in the GBM exome. These include non-silent mutations in genes that have roles in the cancer, including *TP53*, *EGFR*, and *PTEN*. Also, the recurrent somatic copy number alteration events in the cohort exhibit concordance with observations from multiple studies. Surveys of the non-coding changes in the GBM genome have heretofore been sporadically undertaken. These include examining the roles of non-coding RNAs in GBM pathogenesis [238, 239], and analysis of the mutations in the promoter for the *TERT* gene [221]. In the current inquiry, we focused on non-coding changes in the regions of the genome that are evolutionarily constrained, and we discovered that there is a significant enrichment of what we designate as non-coding constraint mutations in the neighborhood of previously implicated GBM genes. A large number of these mutations overlap cis-regulatory elements, and somatic mutations in these elements/features possibly result in modulating/altering the expression of the genes to which they are linked. In addition, we also identify 43 other genes that have NCCMs >3.0/100 kbp sequence, potentially providing additional/novel GBM candidate genes. Overall, we believe harnessing the information garnered with the (potential) driving non-coding alterations from our study, along with the knowledge we have about protein-coding changes, can serve to inform on probable modes of carcinogenesis adopted in glioblastoma.

## 6.2 Limitations of the current investigations

A critical and objective evaluation of our three studies is liable to reveal a few shortcomings, which is crucial to address and resolve for the follow-up investigations that may arise. Across all our studies, we have a few tens of samples, leading to relatively low power, in particular for weaker effects. Though we were able to delineate major candidate drivers for the diseases investigated, there potentially exist other genes, with fewer or weaker alterations that work synergistically to produce effects that have impacts on carcinogenic initiation and/or progression. These are likely to be left undetected or fall short of a statistically significant result when working with non-optimal sample sizes. It

is also possible when working with a (small) study population that ostensibly significant results, despite passing appropriate thresholds for the test statistic used, are in truth false-positives or have overestimated the effect size of the association between genotype and phenotype, a phenomenon known as “Winner’s Curse”.

A second limitation could be a consequence arising from the scope of the study. For instance, in our investigations into osteosarcoma and hemangiosarcoma with WES, we focused on defining only the coding mutational landscape. Any effects exerted by the non-coding genome, as a result, remain unidentified. Another factor scope-wise is the number of breeds chosen for the study. Though we have selected breed(s) that are susceptible to the diseases investigated, it would be beneficial to know if the same disease variants are associated across other diverse populations, or if there are varying breed-specific routes to arriving at the phenotype.

In our investigation of GBM, the whole-genome approach removed the lack of non-coding information described above, which gave us important insights into genes and non-coding alterations that have potential roles in carcinogenesis. However, we need to be able to perform functional assays or characterizations to confirm these somatic alterations are actually involved in initiating cancer or have roles in carcinogenesis. Only when this is established can we propose the discerned somatic changes as clinically actionable genetic events for future therapeutic strategies.

Genetic tumor heterogeneity is another issue to cope with in our studies, and in general, across all cancer sequencing investigations. This phenomenon, stemming from spatial and temporal changes in a given tumor, can give rise to multiple subpopulations of tumor cells or heterogeneous subclones and is observed across most cancers [240]. Bulk sequencing, which is generally employed to profile these tumor subclones, may not capture all of the cells that contain driver alterations, which can lead to a decrease in the sensitivity of mutation detection [241, 242].

Lastly, across all our studies, we have focused on probing the disease at the DNA level. However, the actual disease manifestation is the result of alterations that occur across multiple strata, from the genome, epigenome, transcriptome, proteome, and the metabolome. While there are individual-level –omics approaches to studying each of these layers—for example, epigenomics, to identify altered epigenetic landscapes, or transcriptomics, to discover the differential expression of mRNA—we need to have integrative methodologies to systematically interrogate and comprehend the fundamental mechanisms governing the biology of cancer.

### 6.3 Ongoing work and future directions

Our work thus far—the delineation of somatic alterations in the genomes of the osteosarcoma, hemangiosarcoma, and glioblastoma—is only the first step in comprehensively defining the mutational landscape of these complex cancers. We are involved in several ongoing explorations for each of these diseases.

For osteosarcoma, we are in the process of sequencing more prone-breeds, using both WGS as well as RNA-seq methodologies. The former will likely give us insights into the non-coding genome's role in the cancer, and the latter is likely to increase our knowledge about aberrant genetic alterations and dysregulated molecular pathways in the disease and the relationship between the two. Our quest to expand our understanding of germline susceptibility for the disease as detailed by Karlsson *et al.* [108] is continuing – we are in the process of collecting data for several new breeds of dogs (including golden retrievers, Leonbergers, and Great Danes, among others) to conduct another GWAS study and possibly discover new risk loci and gene pathways that might be implicated for breed-specific OSA. Also, we will select candidate loci from the new research and previously identified 33 loci, and perform fine mapping using first whole-genome sequencing and then targeted genotyping and functional analysis to hone in on potential causative loci. Evidence from several cancers indicates that germline mutations cooperate with somatic mutations to drive carcinogenesis [243-245]. With WGS data from both tumors as well as normal samples available, we plan to conduct a similar exploratory analysis to see if there are molecular links present between germline predisposing factors and the alterations that occur in the somatic genome.

In the study of hemangiosarcoma, we are exploring the possibility of using liquid biopsy, a noninvasive approach to study the molecular profiles of tumors. This procedure involves sequencing circulating tumor DNA (ctDNA) from small volumes of secreted body fluids such as blood, urine, or saliva [246]. Because ctDNA can be a proxy for the tumor genome, the use of liquid biopsies for sequencing may aid in getting an accurate snapshot of the driver mutations that are present in the tumor. Currently, we have tested ctDNA sequencing as a proof of concept for several dogs across multiple breeds, and we find that the approach is compatible for a vascular cancer like HSA. Furthermore, with collaborating veterinarians, we propose to use the liquid biopsy for longitudinal sampling or monitoring of the cancer, *i.e.*, screen for tumor lesions that may arise in predisposing patients over time.

For glioblastoma also, we are expanding our cohort to not only whole-genome sequence data from more patients but in addition are also sequencing patient-derived cell lines corresponding to the matched tumor-normal of the existing data as well as from the above dataset. Our goals here are to make a systematic genomic comparison of the cell line and tumor mutation data and validate to what extent the mutations in cell lines are representative of the



original tumor, and to estimate the extent of heterogeneity in the original tumors compared to the resulting cell lines. We are also engaged in RNA sequencing of the same tumor tissues used in the present study, to monitor gene expression and transcriptome changes in candidate genes, and also for the coupling of tumor NCCMs with expression data.

As alluded to in the previous section, looking at only the mutational changes at the DNA level gives only a partial glimpse of the alterations in the cancer genome. Therefore, an overarching goal for all of the projects here (and across cancer studies in general) would be to collect biological data that encompasses the dimensions of epigenome, transcriptome, proteome, and metabolome, and to use the derived multi-omic measurements to understand the foundations of cancer in terms of biological mechanism, possible biomarkers and, most importantly, therapeutic targets.

## 7 Acknowledgements

गुरु साक्षात परब्रह्मा तस्मै श्रीगुरवे नमः

**Guru Saakshaat Para Brahma Tasmai Sri Gurave Namaha**

The above phrase is an excerpt from one of the doctrines in Indian philosophy. It roughly translates that a ‘guru’ is an individual who is a fount of knowledge and someone who dispels the weeds of ignorance in his or her followers (*i.e.*, students). I have been fortunate to have two such gurus, **Kerstin Lindblad-Toh** and **Karin Forsberg-Nilsson**, guiding me through the course of my PhD.

**Kerstin**, I cannot thank you enough for taking me on as a student. Before joining your group, I knew of your brilliance and ingenuity by reputation, and in the last five years, it has been an absolute privilege to witness these first-hand and have you mentor me through the various projects.

**Karin**, a scientist who effortlessly dons on many hats, thank you for letting me be part of your group and share your vision in the glioblastoma projects. Your constant words of encouragement and support through the years have spurred me to overcome (real and imagined) roadblocks and become a better researcher.

**Voichița Marinescu (Vikki)**, my co-supervisor, despite you telling me at the start of my PhD, that I am going to embark on a solo and arduous journey, you have been there to give guidance, support, and feedback the entire time. And I thank you for that! Also, many thanks for shooting the best-ever images of me!

**Mats Pettersson**, thank you for letting me pick your brilliant mind on many occasions over the years. And thank you for being a sounding board par excellence during the writing of this thesis.

**Maja Arendt** and **Johanna Dahlqvist**, thank you for your invaluable input medical-terminology-wise during my writing. (MA, also thank you and Daniel for putting a roof over my head when I first arrived in Uppsala. JD, it has been an absolute treat to have our customary July/bi-weekly fika and natter sessions.)

To all of the **KLT** and **KFN** group members – thank you for the meetings, JCs, and collaborations. I am smarter and in the know about research and current trends as a result of them.

Thank you, **Markus Mayrhofer** and **Björn Nystedt**, at the SciLifeLab Bioinformatics advisory program for your valuable inputs on all things bioinformatics.

There are several people at the Broad Institute to whom I owe a debt of gratitude: many thanks to **Christina Cuomo** and **Bruce Birren**, for encouraging me to pursue an advanced degree. **Jeremy Johnson** and **Jessica Alföldi** for urging me to seek a position with Kerstin, with the disclaimer – despite being an über PI, KLT does not bite! Thank you, **Kate Megquier**, **Jason Turner-Maier**, and **Ross Swofford** for being an email away from answering queries big and small. **Leslie Gaffney**, thank you for letting me take advantage of our decade-old friendship and have you design (what I think) is a fantastic cover for this thesis.

I will be remiss if I don't acknowledge some of the biggest influences in my scientific career, from a high school teacher to world-renowned researchers. Yes! I am name-dropping, but these are folks for whom science is a verb akin to breathing. The list includes **Sribhaghyam Srinivas**, **Malini Nagaraju**, **Roger W. Read**, **Warren J. Ewens**, **Kerstin Lindblad-Toh**, and **Leif Andersson**.

Teaching, a pedagogical criterion, became a surprisingly pleasant endeavor. This is in no small measure to the following people: **Dorothe Spillmann**, **Eva Gottfridsson**, **Lionel Guy**, **Per Jemth**, **Gerli Rosengren Pielberg**, and **Jessika Nordin**. (JN, I owe you a lot more than what can be spelled here!)

To all present and past fellow D11 corridor friends/researchers/office-mates (a list that is *too* numerous to enumerate), I thank every one of you for all the laughter, fika, food, and tête-à-têtes. If I have offended anybody with my 'no-filters' exchange, I humbly apologize for it here.

**Karthik**, my partner in 'Silent Symphony' (an inside joke dear reader, please move on to the next paragraph), despite or should I say in spite of his (massive) qualms, is my biggest champion. He has consented to a trans-Atlantic relationship, let me navigate my route, and become the person I am today.

**My family – Sakthi Appa, Geetha Amma, Vasu, and Anu** for your unceasing love and encouragement. Also, many thanks to the **Thumbavanam** and the **Palayur** family branches for your support. Your conviction in my abilities has helped me plow through the journey towards a PhD.

Lastly, (and probably the first time ever thanked in a thesis), thank you to the magical **Swedish skog** for giving me endless solace and joy during my stint in the country! Of all things Swedish – I AM GOING TO MISS YOU THE MOST!!

## 8 References

1. Mukherjee, S., *The emperor of all maladies: A biography of cancer*. 2010, New York: Scribner.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**.
3. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation*. Cell, 2011. **144**(5): p. 646-674.
4. Frank, S.A., in *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. 2007, Princeton University Press.
5. Manchester, K.L., *Theodor Boveri and the origin of malignant tumours*. Trends Cell Biol, 1995. **5**(10): p. 384-7.
6. Renda, A., *Multiple Primary Malignancies*. 1st ed. 2009, Naples, Italy: Springer Italy.
7. Nordling, C.O., *A new theory on cancer-inducing mechanism*. Br J Cancer, 1953. **7**(1): p. 68-72.
8. Siddiqui, I.A., et al., *Resveratrol nanoformulation for cancer prevention and therapy*. Ann N Y Acad Sci, 2015. **1348**(1): p. 20-31.
9. Ashley, D.J., *Environmental factors in the aetiology of lung cancer and bronchitis*. Br J Prev Soc Med, 1969. **23**(4): p. 258-62.
10. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma*. Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
11. Chinnam, M. and D.W. Goodrich, *RBI, development, and cancer*. Curr Top Dev Biol, 2011. **94**: p. 129-69.
12. Lee, E.Y. and W.J. Muller, *Oncogenes and tumor suppressor genes*. Cold Spring Harb Perspect Biol, 2010. **2**(10): p. a003236.
13. Weinberg, R.A., *The biology of cancer*. 2014.
14. Olivier, M., et al., *TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer*. IARC Sci Publ, 2004(157): p. 247-70.
15. Duncan, J.A., J.R. Reeves, and T.G. Cooke, *BRCA1 and BRCA2 proteins: roles in health and disease*. Mol Pathol, 1998. **51**(5): p. 237-47.
16. Yang, X. and M.E. Lippman, *BRCA1 and BRCA2 in breast cancer*. Breast Cancer Res Treat, 1999. **54**(1): p. 1-10.
17. Milella, M., et al., *PTEN: Multiple Functions in Human Malignant Tumors*. Frontiers in oncology, 2015. **5**: p. 24-24.
18. Rodenhuis, S., *ras and human tumors*. Semin Cancer Biol, 1992. **3**(4): p. 241-7.
19. Minamoto, T., M. Mai, and Z. Ronai, *K-ras mutation: early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers--a review*. Cancer Detect Prev, 2000. **24**(1): p. 1-12.
20. Padua, R.A. and R.R. West, *Oncogene mutation and prognosis in the myelodysplastic syndromes*. Br J Haematol, 2000. **111**(3): p. 873-4.

21. Chen, H., H. Liu, and G. Qing, *Targeting oncogenic Myc as a strategy for cancer treatment*. Signal Transduct Target Ther, 2018. **3**: p. 5.
22. Huebner, R.J. and G.J. Todaro, *Oncogenes of RNA tumor viruses as determinants of cancer*. Proc Natl Acad Sci U S A, 1969. **64**(3): p. 1087-94.
23. Stehelin, D., et al., *DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA*. Nature, 1976. **260**(5547): p. 170-3.
24. Adams, J.M., et al., *The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice*. Nature, 1985. **318**(6046): p. 533-8.
25. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394-424.
26. World Health Organization. *World Health Organization Cancer Key Facts*. 2017; Available from: <http://www.who.int/mediacentre/factsheets/fs297/en/>.
27. Mariotto, A.B., et al., *Projections of the cost of cancer care in the United States: 2010-2020*. J Natl Cancer Inst, 2011. **103**(2): p. 117-28.
28. Ershler, W.B., *Cancer: a disease of the elderly*. J Support Oncol, 2003. **1**(4 Suppl 2): p. 5-10.
29. Permauer, I. and N. Scholl, *Global trends in lifespan inequality: 1950-2015*. PloS one, 2019. **14**(5): p. e0215742-e0215742.
30. Yabroff, K.R., et al., *Economic burden of cancer in the United States: estimates, projections, and future research*. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2011. **20**(10): p. 2006-2014.
31. International Human Genome Sequencing, C., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**: p. 860.
32. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
33. de Koning, T.J., et al., *Targeted next-generation sequencing panels for monogenetic disorders in clinical diagnostics: the opportunities and challenges*. Expert Rev Mol Diagn, 2015. **15**(1): p. 61-70.
34. Lim, E.C., et al., *Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients*. Hum Genomics, 2015. **9**: p. 33.
35. Shen, T., et al., *Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes*. Front Genet, 2015. **6**: p. 215.
36. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation*. American journal of human genetics, 2017. **101**(1): p. 5-22.
37. Boycott, K.M., et al., *Rare-disease genetics in the era of next-generation sequencing: discovery to translation*. Nat Rev Genet, 2013. **14**(10): p. 681-91.
38. Keogh, M.J. and P.F. Chinnery, *Next generation sequencing for neurological diseases: New hope or new hype?* Clinical Neurology and Neurosurgery, 2013. **115**(7): p. 948-953.
39. Ma, Y., et al., *Applications of Next-generation Sequencing in Systemic Autoimmune Diseases*. Genomics, Proteomics & Bioinformatics, 2015. **13**(4): p. 242-249.

40. Precone, V., et al., *Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives*. BioMed research international, 2015. **2015**: p. 161648-161648.
41. Bush, W.S. and J.H. Moore, *Chapter 11: Genome-wide association studies*. PLoS Comput Biol, 2012. **8**(12): p. e1002822.
42. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations*. Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
43. Shyr, D. and Q. Liu, *Next generation sequencing in cancer research and clinical application*. Biological Procedures Online, 2013. **15**(1): p. 4.
44. Heng, J., et al., *Integrated analysis of promoter mutation, methylation and expression of AKT1 gene in Chinese breast cancer patients*. PLoS One, 2017. **12**(3): p. e0174022.
45. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy*. Science, 2004. **304**(5676): p. 1497-500.
46. Wooster, R., A.P. Futreal, and M.R. Stratton, *Sequencing analysis of BRAF mutations in human cancers*. Methods Enzymol, 2006. **407**: p. 218-24.
47. Shendure, J. and H. Ji, *Next-generation DNA sequencing*. Nat Biotechnol, 2008. **26**(10): p. 1135-45.
48. Muzzey, D., E.A. Evans, and C. Lieber, *Understanding the Basics of NGS: From Mechanism to Variant Calling*. Current genetic medicine reports, 2015. **3**(4): p. 158-165.
49. Metzker, M.L., *Sequencing technologies - the next generation*. Nat Rev Genet, 2010. **11**(1): p. 31-46.
50. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies*. Nat Rev Genet, 2016. **17**(6): p. 333-51.
51. Koboldt, D.C., et al., *The next-generation sequencing revolution and its impact on genomics*. Cell, 2013. **155**(1): p. 27-38.
52. Meyerson, M., S. Gabriel, and G. Getz, *Advances in understanding cancer genomes through second-generation sequencing*. Nat Rev Genet, 2010. **11**(10): p. 685-96.
53. Goh, L., et al., *Assessing matched normal and tumor pairs in next-generation sequencing studies*. PLoS One, 2011. **6**(3): p. e17810.
54. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry*. Nature, 2008. **456**(7218): p. 53-9.
55. Wilmott, J.S., et al., *Tumour procurement, DNA extraction, coverage analysis and optimisation of mutation-detection algorithms for human melanoma genomes*. Pathology, 2015. **47**(7): p. 683-93.
56. Brastianos, P.K., et al., *Genomic sequencing of meningiomas identifies oncogenic SMO and AKT1 mutations*. Nature genetics, 2013. **45**(3): p. 285-289.
57. Desmedt, C., et al., *Next-generation sequencing in breast cancer: first take home messages*. Current opinion in oncology, 2012. **24**(6): p. 597-604.
58. Coco, S., et al., *Next generation sequencing in non-small cell lung cancer: new avenues toward the personalized medicine*. Curr Drug Targets, 2015. **16**(1): p. 47-59.
59. Ojesina, A.I., et al., *Landscape of genomic alterations in cervical carcinomas*. Nature, 2014. **506**(7488): p. 371-5.

60. Han, S.W., et al., *Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing*. PLoS One, 2013. **8**(5): p. e64271.
61. Chapman, M.A., et al., *Initial genome sequencing and analysis of multiple myeloma*. Nature, 2011. **471**(7339): p. 467-472.
62. Sakamoto, Y., S. Sereewattanawoot, and A. Suzuki, *A new era of long-read sequencing for cancer genomics*. Journal of Human Genetics, 2019.
63. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome*. Nature, 2002. **420**(6915): p. 520-62.
64. Lindblad-Toh, K., et al., *Genome sequence, comparative analysis and haplotype structure of the domestic dog*. Nature, 2005. **438**(7069): p. 803-19.
65. Hughes, A.L., *Genomes of mice and men*. Heredity (Edinb), 2003. **90**(2): p. 115-7.
66. Arendt, M.L., et al., *Genome-Wide Association Study of Golden Retrievers Identifies Germ-Line Risk Factors Predisposing to Mast Cell Tumours*. PLoS Genet, 2015. **11**(11): p. e1005647.
67. Starkey, M.P., et al., *Dogs really are man's best friend--canine genomics has applications in veterinary and human medicine!* Brief Funct Genomic Proteomic, 2005. **4**(2): p. 112-28.
68. Sargan, D.R., *IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics*. Mamm Genome, 2004. **15**(6): p. 503-6.
69. Parker, H.G., A.L. Shearin, and E.A. Ostrander, *Man's best friend becomes biology's best in show: genome analyses in the domestic dog*. Annu Rev Genet, 2010. **44**: p. 309-36.
70. Ostrander, E.A., *Unleashing the Canine Genome*. Genome Research, 2000. **10**(9): p. 1271-1274.
71. Ostrander, E.A., F. Galibert, and D.F. Patterson, *Canine genetics comes of age*. Trends Genet, 2000. **16**(3): p. 117-24.
72. Morello, E., M. Martano, and P. Buracco, *Biology, diagnosis and treatment of canine appendicular osteosarcoma: similarities and differences with human osteosarcoma*. Vet J, 2011. **189**(3): p. 268-77.
73. Ito, D., A.M. Frantz, and J.F. Modiano, *Canine lymphoma as a comparative model for human non-Hodgkin lymphoma: recent progress and applications*. Vet Immunol Immunopathol, 2014. **159**(3-4): p. 192-201.
74. Queiroga, F.L., et al., *Canine mammary tumours as a model to study human breast cancer: most recent findings*. In Vivo, 2011. **25**(3): p. 455-65.
75. Dickinson, P.J., et al., *Chromosomal Aberrations in Canine Gliomas Define Candidate Genes and Common Pathways in Dogs and Humans*. Journal of Neuropathology & Experimental Neurology, 2016.
76. Mohammed, S.I., et al., *Expression of cyclooxygenase-2 (COX-2) in human invasive transitional cell carcinoma (TCC) of the urinary bladder*. Cancer Res, 1999. **59**(22): p. 5647-50.
77. Khanna, C., et al., *The dog as a cancer model*. Nat Biotech, 2006. **24**(9): p. 1065-1066.
78. Tang, J., et al., *Copy number abnormalities in sporadic canine colorectal cancers*. Genome Res, 2010. **20**(3): p. 341-50.
79. Elvers, I., et al., *Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background*. Genome Res, 2015. **25**(11): p. 1634-45.

80. Schiffman, J.D. and M. Breen, *Comparative oncology: what dogs and other species can teach us about humans with cancer*. Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1673).
81. Tonomura, N., et al., *Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers*. PLoS Genet, 2015. **11**(2): p. e1004922.
82. Moreno, C., et al., *Comparative genomics for detecting human disease genes*. Adv Genet, 2008. **60**: p. 655-97.
83. Clutton-Brock, J., *A Natural History of Domesticated Mammals*. 2nd ed. 1999: Cambridge University Press.
84. Karlsson, E.K. and K. Lindblad-Toh, *Leader of the pack: gene mapping in dogs and other model organisms*. Nat Rev Genet, 2008. **9**(9): p. 713-25.
85. Ostrander, E.A., *Franklin H. Epstein Lecture. Both ends of the leash--the human links to good dogs with bad genes*. N Engl J Med, 2012. **367**(7): p. 636-46.
86. Nobrega, M.A. and L.A. Pennacchio, *Comparative genomic analysis as a tool for biological discovery*. The Journal of physiology, 2004. **554**(Pt 1): p. 31-39.
87. Cooper, G.M. and C.D. Brown, *Qualifying the relationship between sequence conservation and molecular function*. Genome Res, 2008. **18**(2): p. 201-5.
88. Hrdlickova, B., et al., *Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease*. Biochim Biophys Acta, 2014. **1842**(10): p. 1910-1922.
89. Davis, C.A., et al., *The Encyclopedia of DNA elements (ENCODE): data portal update*. Nucleic Acids Res, 2018. **46**(D1): p. D794-d801.
90. Lindblad-Toh, K., et al., *A high-resolution map of human evolutionary constraint using 29 mammals*. Nature, 2011. **478**(7370): p. 476-82.
91. Davydov, E.V., et al., *Identifying a high fraction of the human genome to be under selective constraint using GERP++*. PLoS Comput Biol, 2010. **6**(12): p. e1001025.
92. Jo, V.Y. and C.D. Fletcher, *WHO classification of soft tissue tumours: an update based on the 2013 (4th) edition*. Pathology, 2014. **46**(2): p. 95-104.
93. Ottaviani, G. and N. Jaffe, *The epidemiology of osteosarcoma*. Cancer Treat Res, 2009. **152**: p. 3-13.
94. Daw, N.C., et al., *Metastatic osteosarcoma*. Cancer, 2006. **106**(2): p. 403-12.
95. Lau, C.C., et al., *Frequent amplification and rearrangement of chromosomal bands 6p12-p21 and 17p11.2 in osteosarcoma*. Genes Chromosomes Cancer, 2004. **39**(1): p. 11-21.
96. Kuijjer, M.L., P.C. Hogendoorn, and A.M. Cleton-Jansen, *Genome-wide analyses on high-grade osteosarcoma: making sense of a genomically most unstable tumor*. Int J Cancer, 2013. **133**(11): p. 2512-21.
97. Chen, X., et al., *Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma*. Cell Rep, 2014. **7**(1): p. 104-12.
98. Wong, F.L., et al., *Cancer incidence after retinoblastoma: radiation dose and sarcoma risk*. Jama, 1997. **278**(15): p. 1262-1267.
99. Jundt, G. and D. Baumhoer, *Hereditary bone tumors*. Pathologe, 2010. **31**(6): p. 471-6.



100. Ehrhart, N.P., S.D. Ryan, and T.M. Fan, *Withrow and MacEwen's Small Animal Clinical Oncology, Ch. 24 - Tumors of the Skeletal System*, in *Withrow and MacEwen's Small Animal Clinical Oncology (Fifth Edition)*, D.M. Vail and R.L. Page, Editors. 2013, W.B. Saunders: Saint Louis. p. 463-503.
101. Rowell, J.L., D.O. McCarthy, and C.E. Alvarez, *Dog Models of Naturally Occurring Cancer*. Trends in molecular medicine, 2011. **17**(7): p. 380-388.
102. Goldschmidt, M.H. and D.E. Thrall, *Malignant bone tumors in the dog*. NEWTON CD; NUNAMAKER, DM Textbook of Small Animal Orthopaedics. Ithaca: International Veterinary Information Service, 1985.
103. Makielski, K.M., et al., *Risk Factors for Development of Canine and Human Osteosarcoma: A Comparative Review*. Veterinary sciences, 2019. **6**(2): p. 48.
104. Banks, W.C., *Parosteal osteosarcoma in a dog and a cat*. J Am Vet Med Assoc, 1971. **158**(8): p. 1412-5.
105. Simpson, S., et al., *Comparative review of human and canine osteosarcoma: morphology, epidemiology, prognosis, treatment and genetics*. Acta veterinaria Scandinavica, 2017. **59**(1): p. 71-71.
106. Cadieu, E. and E.A. Ostrander, *Canine genetics offers new mechanisms for the study of human cancer*. Cancer Epidemiol Biomarkers Prev, 2007. **16**(11): p. 2181-3.
107. Egenvall, A., A. Nødtvedt, and H. von Euler, *Bone tumors in a population of 400 000 insured Swedish dogs up to 10 y of age: incidence and survival*. Canadian Journal of Veterinary Research, 2007. **71**(4): p. 292-299.
108. Karlsson, et al., *Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B*. Genome Biol, 2013. **14**(12): p. R132.
109. Martin, J.W., J.A. Squire, and M. Zielenska, *The genetics of osteosarcoma*. Sarcoma, 2012. **2012**: p. 627254.
110. Meyers, P.A., et al., *Osteosarcoma: a randomized, prospective trial of the addition of ifosfamide and/or muramyl tripeptide to cisplatin, doxorubicin, and high-dose methotrexate*. J Clin Oncol, 2005. **23**(9): p. 2004-11.
111. Fenger, J.M., C.A. London, and W.C. Kisseberth, *Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology*. ILAR J, 2014. **55**(1): p. 69-85.
112. Thanindrarn, P., et al., *Advances in immune checkpoint inhibitors for bone sarcoma therapy*. Journal of bone oncology, 2019. **15**: p. 100221-100221.
113. Lamerato-Kozicki, A.R., et al., *Canine hemangiosarcoma originates from hematopoietic precursors with potential for endothelial differentiation*. Exp Hematol, 2006. **34**(7): p. 870-8.
114. Kim, J.H., et al., *Pathobiology of Hemangiosarcoma in Dogs: Research Advances and Future Perspectives*. Vet Sci, 2015. **2**(4): p. 388-405.
115. Chaikin, P. and A. Welihozkiy, *Hemangiosarcoma in a Dog: Unusual Presentation and Increased Survival Using a Complementary/Holistic Approach Combined with Metronomic Chemotherapy*. Case Rep Vet Med, 2018. **2018**: p. 6160980.
116. Sorenmo, K.U., et al., *Efficacy and toxicity of a dose-intensified doxorubicin protocol in canine hemangiosarcoma*. J Vet Intern Med, 2004. **18**(2): p. 209-13.

117. Tamburini, B.A., et al., *Gene expression profiling identifies inflammation and angiogenesis as distinguishing features of canine hemangiosarcoma*. BMC Cancer, 2010. **10**: p. 619.
118. Wang, G., et al., *Actionable mutations in canine hemangiosarcoma*. PLoS One, 2017. **12**(11): p. e0188667.
119. Young, R.J., et al., *Angiosarcoma*. Lancet Oncol, 2010. **11**(10): p. 983-91.
120. Cioffi, A., et al., *Angiosarcomas and other sarcomas of endothelial origin*. Hematol Oncol Clin North Am, 2013. **27**(5): p. 975-88.
121. Yin, M., et al., *Prognosis and treatment of non-metastatic primary and secondary breast angiosarcoma: a comparative study*. BMC cancer, 2017. **17**(1): p. 295-295.
122. Macmahon, H.E., A.S. Murphy, and M.I. Bates, *Endothelial-Cell Sarcoma of Liver Following Thorotrast Injections*. Am J Pathol, 1947. **23**(4): p. 585-611.
123. Elliott, P. and I. Kleinschmidt, *Angiosarcoma of the liver in Great Britain in proximity to vinyl chloride sites*. Occup Environ Med, 1997. **54**(1): p. 14-8.
124. Marion, M.J. and S. Boivin-Angele, *Vinyl chloride-specific mutations in humans and animals*. IARC Sci Publ, 1999(150): p. 315-24.
125. Manner, J., et al., *MYC high level gene amplification is a distinctive feature of angiosarcomas after irradiation or chronic lymphedema*. Am J Pathol, 2010. **176**(1): p. 34-9.
126. Guo, T., et al., *Consistent MYC and FLT4 gene amplification in radiation-induced angiosarcoma but not in other radiation-associated atypical vascular lesions*. Genes Chromosomes Cancer, 2011. **50**(1): p. 25-33.
127. Antonescu, C.R., et al., *KDR activating mutations in human angiosarcomas are sensitive to specific kinase inhibitors*. Cancer Res, 2009. **69**(18): p. 7175-9.
128. Behjati, S., et al., *Recurrent PTPRB and PLCG1 mutations in angiosarcoma*. Nat Genet, 2014. **46**(4): p. 376-379.
129. Singla, S., et al., *Challenges in the treatment of angiosarcoma: a single institution experience*. Am J Surg, 2014. **208**(2): p. 254-9.
130. Gustafson, D.L., et al., *Canine sarcomas as a surrogate for the human disease*. Pharmacology & therapeutics, 2018. **188**: p. 80-96.
131. Weller, M., et al., *Glioma*. Nature Reviews Disease Primers, 2015. **1**(1): p. 15017.
132. Collins, V.P., D.T.W. Jones, and C. Giannini, *Pilocytic astrocytoma: pathology, molecular mechanisms and markers*. Acta neuropathologica, 2015. **129**(6): p. 775-788.
133. Holland, E.C., *Glioblastoma multiforme: the terminator*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6242-6244.
134. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary*. Acta Neuropathol, 2016. **131**(6): p. 803-20.
135. Turkalp, Z., J. Karamchandani, and S. Das, *IDH mutation in glioma: new insights and promises for the future*. JAMA Neurol, 2014. **71**(10): p. 1319-25.

136. Chen, J.-R., et al., *Isocitrate Dehydrogenase (IDH)1/2 Mutations as Prognostic Markers in Patients With Glioblastomas*. *Medicine*, 2016. **95**(9): p. e2583-e2583.
137. Ohgaki, H. and P. Kleihues, *The Definition of Primary and Secondary Glioblastoma*. *Clinical Cancer Research*, 2013. **19**(4): p. 764.
138. Melin, B.S., et al., *Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors*. *Nature Genetics*, 2017. **49**: p. 789.
139. Karlbom, A.E., et al., *Loss of heterozygosity in malignant gliomas involves at least three distinct regions on chromosome 10*. *Human Genetics*, 1993. **92**(2): p. 169-174.
140. Albarosa, R., et al., *Deletion mapping of gliomas suggest the presence of two small regions for candidate tumor-suppressor genes in a 17-cM interval on chromosome 10q*. *American journal of human genetics*, 1996. **58**(6): p. 1260-1267.
141. Maier, D., et al., *New deletion in low-grade oligodendroglioma at the glioblastoma suppressor locus on chromosome 10q25-26*. *Oncogene*, 1997. **15**(8): p. 997-1000.
142. Ichimura, K., et al., *Distinct patterns of deletion on 10p and 10q suggest involvement of multiple tumor suppressor genes in the development of astrocytic gliomas of different malignancy grades*. *Genes Chromosomes Cancer*, 1998. **22**(1): p. 9-15.
143. Gelpi, E., et al., *Fluorescent in situ hybridization on isolated tumor cell nuclei: a sensitive method for 1p and 19q deletion analysis in paraffin-embedded oligodendroglial tumor specimens*. *Mod Pathol*, 2003. **16**(7): p. 708-15.
144. Jenkins, R.B., et al., *A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma*. *Cancer Res*, 2006. **66**(20): p. 9852-61.
145. Tews, B., et al., *Identification of novel oligodendroglioma-associated candidate tumor suppressor genes in 1p36 and 19q13 using microarray-based expression profiling*. *Int J Cancer*, 2006. **119**(4): p. 792-800.
146. Eckel-Passow, J.E., et al., *Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors*. *The New England journal of medicine*, 2015. **372**(26): p. 2499-2508.
147. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma*. *Cell*, 2013. **155**(2): p. 462-77.
148. Friedmann-Morvinski, D., *Glioblastoma heterogeneity and cancer cell plasticity*. *Crit Rev Oncog*, 2014. **19**(5): p. 327-36.
149. Inda, M.-d.-M., R. Bonavia, and J. Seoane, *Glioblastoma Multiforme: A Look Inside Its Heterogeneous Nature*. *Cancers*, 2014. **6**(1): p. 226-239.
150. Sottoriva, A., et al., *Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics*. *Proc Natl Acad Sci U S A*, 2013. **110**(10): p. 4009-14.
151. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma*. *Science*, 2014. **344**(6190): p. 1396-401.
152. Hottinger, A.F., R. Stupp, and K. Homicsko, *Standards of care and novel approaches in the management of glioblastoma multiforme*. *Chin J Cancer*, 2014. **33**(1): p. 32-9.

153. Chiocca, E.A., et al., *Regulatable interleukin-12 gene therapy in patients with recurrent high-grade glioma: Results of a phase I trial*. Sci Transl Med, 2019. **11**(505).
154. Lim, M., et al., *Current state of immunotherapy for glioblastoma*. Nature Reviews Clinical Oncology, 2018. **15**(7): p. 422-442.
155. Snyder, J.M., et al., *Canine intracranial primary neoplasia: 173 cases (1986-2003)*. J Vet Intern Med, 2006. **20**(3): p. 669-75.
156. Song, R.B., et al., *Postmortem evaluation of 435 cases of intracranial neoplasia in dogs and relationship of neoplasm with breed, age, and body weight*. J Vet Intern Med, 2013. **27**(5): p. 1143-52.
157. Truve, K., et al., *Utilizing the Dog Genome in the Search for Novel Candidate Genes Involved in Glioma Development-Genome Wide Association Mapping followed by Targeted Massive Parallel Sequencing Identifies a Strongly Associated Locus*. PLoS Genet, 2016. **12**(5): p. e1006000.
158. Patel, M.B. and J. Wang, *The Identification and Interpretation of cis-Regulatory Noncoding Mutations in Cancer*. High Throughput, 2018. **8**(1).
159. Macconail, L.E. and L.A. Garraway, *Clinical implications of the cancer genome*. J Clin Oncol, 2010. **28**(35): p. 5219-28.
160. Kim, T.M., et al., *Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers*. Clin Cancer Res, 2005. **11**(23): p. 8235-42.
161. Creixell, P., et al., *Pathway and network analysis of cancer genomes*. Nat Methods, 2015. **12**(7): p. 615-621.
162. Wang, E., *Understanding genomic alterations in cancer genomes using an integrative network approach*. Cancer Lett, 2013. **340**(2): p. 261-9.
163. Robinson, P.N., P. Krawitz, and S. Mundlos, *Strategies for exome and genome sequence data analysis in disease-gene discovery projects*. Clin Genet, 2011. **80**(2): p. 127-32.
164. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes*. Nature, 2009. **461**(7261): p. 272-6.
165. Yang, Y., et al., *Molecular findings among patients referred for clinical whole-exome sequencing*. JAMA, 2014. **312**(18): p. 1870-1879.
166. Baldrige, D., et al., *The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results*. Genet Med, 2017. **19**(9): p. 1040-1048.
167. Liang, H., et al., *Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer*. Genome Res, 2012. **22**(11): p. 2120-9.
168. Yap, K.L., et al., *Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival*. Clin Cancer Res, 2014. **20**(24): p. 6605-17.
169. Silvestri, V., et al., *Whole-exome sequencing and targeted gene sequencing provide insights into the role of PALB2 as a male breast cancer susceptibility gene*. Cancer, 2017. **123**(2): p. 210-218.
170. Zare, F., et al., *An evaluation of copy number variation detection tools for cancer using whole exome sequencing data*. BMC Bioinformatics, 2017. **18**(1): p. 286.

171. Meienberg, J., et al., *Clinical sequencing: is WGS the better WES?* Human genetics, 2016. **135**(3): p. 359-362.
172. Khurana, E., et al., *Role of non-coding sequence variants in cancer.* Nat Rev Genet, 2016. **17**(2): p. 93-108.
173. Heidenreich, B., et al., *TERT promoter mutations in cancer development.* Curr Opin Genet Dev, 2014. **24**: p. 30-7.
174. Huarte, M., *The emerging role of lncRNAs in cancer.* Nature Medicine, 2015. **21**(11): p. 1253-1261.
175. Bhan, A., M. Soleimani, and S.S. Mandal, *Long Noncoding RNA and Cancer: A New Paradigm.* Cancer Res, 2017. **77**(15): p. 3965-3981.
176. Balas, M.M. and A.M. Johnson, *Exploring the mechanisms behind long noncoding RNAs and cancer.* Noncoding RNA Res, 2018. **3**(3): p. 108-117.
177. Li, X., H. Zhang, and X. Wu, *Long noncoding RNA DLX6-AS1 accelerates the glioma carcinogenesis by competing endogenous sponging miR-197-5p to relieve E2F1.* Gene, 2019. **686**: p. 1-7.
178. Nakagawa, H. and M. Fujita, *Whole genome sequencing analysis for cancer genomics and precision medicine.* Cancer science, 2018. **109**(3): p. 513-522.
179. Rusch, M., et al., *Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome.* Nature Communications, 2018. **9**(1): p. 3962.
180. Schwarze, K., et al., *Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature.* Genetics in Medicine, 2018. **20**(10): p. 1122-1130.
181. Hoeppner, M.P., et al., *An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts.* PLoS One, 2014. **9**(3): p. e91172.
182. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
183. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.
184. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.* Bioinformatics, 2012. **28**(14): p. 1811-7.
185. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
186. Ameer, A., et al., *SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population.* European Journal Of Human Genetics, 2017. **25**: p. 1253.
187. 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 (Austin), 2012. **6**(2): p. 80-92.
188. Ramos, A.H., et al., *Oncotator: cancer variant annotation tool.* Hum Mutat, 2015. **36**(4): p. E2423-9.
189. Dees, N.D., et al., *MuSiC: Identifying mutational significance in cancer genomes.* Genome Research, 2012. **22**(8): p. 1589-1598.
190. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer genes.* Nature, 2013. **499**(7457): p. 214-218.
191. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biology, 2016. **17**(1): p. 122.

192. Casper, J., et al., *The UCSC Genome Browser database: 2018 update*. Nucleic Acids Res, 2018. **46**(D1): p. D762-d769.
193. Thomas-Chollier, M., et al., *Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs*. Nat Protoc, 2011. **6**(12): p. 1860-9.
194. Kim, J., et al., *Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors*. Nat Genet, 2016. **48**(6): p. 600-606.
195. Lee, J., et al., *Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures*. Nucleic Acids Res, 2018. **46**(W1): p. W102-w108.
196. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer*. Nature, 2013. **500**(7463): p. 415-21.
197. Koboldt, D.C., et al., *VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing*. Genome Research, 2012. **22**(3): p. 568-576.
198. Raine, K.M., et al., *ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data*. Curr Protoc Bioinformatics, 2016. **56**: p. 15.9.1-15.9.17.
199. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers*. Genome Biology, 2011. **12**(4): p. 1-14.
200. Sherman, B.T., et al., *DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis*. BMC Bioinformatics, 2007. **8**: p. 426-426.
201. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. BMC Bioinformatics, 2013. **14**(1): p. 1-14.
202. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
203. Duns, G., et al., *Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma*. Cancer Res, 2010. **70**(11): p. 4287-91.
204. Newbold, R.F. and K. Mokbel, *Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis*. Anticancer Res, 2010. **30**(9): p. 3309-11.
205. Li, J., et al., *SETD2: an epigenetic modifier with tumor suppressor functionality*. Oncotarget, 2016. **7**(31): p. 50719-50734.
206. Sawan, C. and Z. Herceg, *Histone modifications and cancer*. Adv Genet, 2010. **70**: p. 57-85.
207. Mottamal, M., et al., *Histone deacetylase inhibitors in clinical studies as templates for new anticancer agents*. Molecules, 2015. **20**(3): p. 3898-941.
208. Chen, X., et al., *Recurrent Somatic Structural Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma*. Cell Reports, 2014. **7**(1): p. 104-112.
209. Lawrence, M.S., et al., *Discovery and saturation analysis of cancer genes across 21 tumour types*. Nature, 2014. **505**(7484): p. 495-501.
210. Thibodeau, B.J., et al., *Mutational landscape of radiation-associated angiosarcoma of the breast*. Oncotarget, 2018. **9**(11): p. 10042-10053.

211. Je, E.M., et al., *Mutational analysis of PIK3CA, JAK2, BRAF, FOXL2, IDH1, AKT1 and EZH2 oncogenes in sarcomas*. *Apmis*, 2012. **120**(8): p. 635-9.
212. Cully, M., et al., *Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis*. *Nat Rev Cancer*, 2006. **6**(3): p. 184-92.
213. Fruman, D.A., et al., *The PI3K Pathway in Human Disease*. *Cell*, 2017. **170**(4): p. 605-635.
214. Lyles, S.E., et al., *In vitro effects of the tyrosine kinase inhibitor, masitinib mesylate, on canine hemangiosarcoma cell lines*. *Vet Comp Oncol*, 2012. **10**(3): p. 223-35.
215. Gardner, H.L., et al., *Maintenance therapy with toceranib following doxorubicin-based chemotherapy for canine splenic hemangiosarcoma*. *BMC Vet Res*, 2015. **11**: p. 131.
216. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers*. *Nature*, 2010. **463**(7283): p. 899-905.
217. Freire, P., et al., *Exploratory analysis of the copy number alterations in glioblastoma multiforme*. *PloS one*, 2008. **3**(12): p. e4076-e4076.
218. Grossman, R.L., et al., *Toward a Shared Vision for Cancer Genomic Data*. *N Engl J Med*, 2016. **375**(12): p. 1109-12.
219. Fan, Y., et al., *Telomerase Expression by Aberrant Methylation of the TERT Promoter in Melanoma Arising in Giant Congenital Nevi*. *The Journal of investigative dermatology*, 2016. **136**(1): p. 339-342.
220. Borah, S., et al., *Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer*. *Science (New York, N.Y.)*, 2015. **347**(6225): p. 1006-1010.
221. Heidenreich, B., et al., *TERT promoter mutations and telomere length in adult malignant gliomas and recurrences*. *Oncotarget*, 2015. **6**(12): p. 10617-10633.
222. Hu, B., et al., *Epigenetic Activation of WNT5A Drives Glioblastoma Stem Cell Differentiation and Invasive Growth*. *Cell*, 2016. **167**(5): p. 1281-1295.e18.
223. Koonin, E.V. and Y.I. Wolf, *Constraints and plasticity in genome and molecular-phenome evolution*. *Nature Reviews Genetics*, 2010. **11**: p. 487.
224. Hui, D.H.F., et al., *Semaphorin 3C as a Therapeutic Target in Prostate and Other Cancers*. *Int J Mol Sci*, 2019. **20**(3).
225. Hao, J. and J.S. Yu, *Semaphorin 3C and Its Receptors in Cancer and Cancer Stem-Like Cells*. *Biomedicines*, 2018. **6**(2).
226. Kaiser, V.B., M.S. Taylor, and C.A. Semple, *Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types*. *PLoS genetics*, 2016. **12**(8): p. e1006207-e1006207.
227. Lentjes, M.H., et al., *The emerging role of GATA transcription factors in development and disease*. *Expert Rev Mol Med*, 2016. **18**: p. e3.
228. Zheng, R. and G.A. Blobel, *GATA Transcription Factors and Cancer*. *Genes Cancer*, 2010. **1**(12): p. 1178-88.
229. Wang, Z., et al., *GATA2 promotes glioma progression through EGFR/ERK/Elk-1 pathway*. *Med Oncol*, 2015. **32**(4): p. 87.
230. Levine, A.J., *p53, the cellular gatekeeper for growth and division*. *Cell*, 1997. **88**(3): p. 323-31.

231. Olivier, M., M. Hollstein, and P. Hainaut, *TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use*. Cold Spring Harbor Perspectives in Biology, 2010. **2**(1): p. a001008.
232. Bouaoun, L., et al., *TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data*. Hum Mutat, 2016. **37**(9): p. 865-76.
233. Mar, B.G., et al., *SETD2 alterations impair DNA damage recognition and lead to resistance to chemotherapy in leukemia*. Blood, 2017.
234. Kanu, N., et al., *SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair*. Oncogene, 2015. **34**(46): p. 5699-708.
235. Jiang, C., et al., *Histone methyltransferase SETD2 regulates osteosarcoma cell growth and chemosensitivity by suppressing Wnt/beta-catenin signaling*. Biochem Biophys Res Commun, 2018. **502**(3): p. 382-388.
236. Gardner, H.L., et al., *Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2*. Commun Biol, 2019. **2**: p. 266.
237. Siegel, R., D. Naishadham, and A. Jemal, *Cancer statistics, 2013*. CA Cancer J Clin, 2013. **63**(1): p. 11-30.
238. Zhang, X.Q. and G.K. Leung, *Long non-coding RNAs in glioma: functional roles and clinical perspectives*. Neurochem Int, 2014. **77**: p. 78-85.
239. Rynkeviciene, R., et al., *Non-Coding RNAs in Glioma*. Cancers, 2018. **11**(1): p. 17.
240. Burrell, R.A., et al., *The causes and consequences of genetic heterogeneity in cancer evolution*. Nature, 2013. **501**(7467): p. 338-45.
241. Rubben, A. and A. Araujo, *Cancer heterogeneity: converting a limitation into a source of biologic information*. J Transl Med, 2017. **15**(1): p. 190.
242. Turajlic, S., et al., *Resolving genetic heterogeneity in cancer*. Nature reviews. Genetics, 2019. **20**(7): p. 404-416.
243. Kanchi, K.L., et al., *Integrated Analysis of Germline and Somatic Variants in Ovarian Cancer*. Nature communications, 2014. **5**: p. 3156-3156.
244. Mamidi, T.K.K., J. Wu, and C. Hicks, *Interactions between Germline and Somatic Mutated Genes in Aggressive Prostate Cancer*. Prostate Cancer, 2019. **2019**: p. 4047680.
245. Wu, J., et al., *Integrating Germline and Somatic Mutation Information for the Discovery of Biomarkers in Triple-Negative Breast Cancer*. International journal of environmental research and public health, 2019. **16**(6): p. 1055.
246. Heitzer, E., P. Ulz, and J.B. Geigl, *Circulating tumor DNA as a liquid biopsy for cancer*. Clin Chem, 2015. **61**(1): p. 112-23.





# Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Medicine 1624*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2020

Distribution: [publications.uu.se](http://publications.uu.se)  
urn:nbn:se:uu:diva-399978